

Marine picoplankton size distribution and optical property contrasts throughout the Atlantic Ocean revealed using flow cytometry

T. J. SMYTH,*  G. A. TARRAN, AND S. SATHYENDRANATH

Plymouth Marine Laboratory, Prospect Place, Plymouth, Devon PL1 3DH, UK

*Corresponding author: tjsm@pml.ac.uk

Received 24 June 2019; accepted 5 October 2019; posted 9 October 2019 (Doc. ID 370578); published 7 November 2019

Depth-resolved flow cytometric observations have been used to determine the size distribution and refractive index (RI) of picoplankton throughout the Atlantic Ocean. *Prochlorococcus* frequently showed double size distribution peaks centered on 0.75 ± 0.25 and 1.75 ± 0.25 μm ; the smallest peak diameters were ≤ 0.65 μm in the equatorial upwelling with larger cells (~ 0.95 μm) in the surface layers of the tropical gyres. *Synechococcus* was strongly monodispersed: the smallest (~ 1.5 μm) and largest cells (~ 2.25 – 2.50 μm) were encountered in the lowest and highest abundance regions, respectively. Typical RI for *Prochlorococcus* was found to be ~ 1.06 , whereas for *Synechococcus* surface RI varied between 1.04–1.08 at high and low abundances, respectively.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

<https://doi.org/10.1364/AO.58.008802>

1. INTRODUCTION

The size distribution and structural composition of phytoplankton cells are not only of fundamental importance to the ecology, biogeochemistry, and optics of the oceans; they are also of critical importance to the tools we use to observe the oceans at large scales, namely, remote sensing (Earth observation), and what we use to predict what will happen to the oceans in the future, namely, ecosystem modeling.

The propagation of light through the ocean water column is governed by its inherent optical properties (IOPs) [1,2], specifically the absorption and scattering characteristics of the seawater and the various substances—particulate or dissolved—it contains [3]. Phytoplankton are a key component in determining this, and therefore information is required, over a wide range of conditions and species, on the IOP-controlling factors of cell size, abundance, and optical properties.

Optical microscopy is useful for determining phytoplankton taxonomy [4], but it is labor intensive and requires many years of specialized training. It is also a low-throughput technique such that single samples of seawater may take several hours to analyze. However, it can render information on species abundance, size, morphology, and, using assumptions, carbon content [5]. In recent years, high-throughput techniques such as FlowCam [6,7] have provided a fast, accurate, easy-to-use alternative to manual microscopy for monitoring plankton community composition. However, for the smaller picoplankton size classes (< 3 μm ; [8]), which cover vast tracts of the global ocean [9],

flow cytometry (FC) arguably provides the best option in terms of its speed, resolution, and ability to enable discrimination of multiple populations within a community [10].

FC measures the light scatter (forward and side) and fluorescence characteristics of a population of cells or particles on an individual basis by injecting fluid (seawater)-suspended samples, ideally flowing one cell at a time, through a laser beam. A combination of fluorescence (typically green, orange, and red) and side scatter properties are used to further sort the phytoplankton cells into different types and hence enumerate abundances [10]. Additional information on the cellular size and refractive index (RI) is retrievable using the forward and side scattering measurements [11–14], provided that the flow cytometer's internal geometry is well characterized and the measurements are frequently referenced against particles of known size and RI [12].

This work makes the following significant advances:

- (1) Finer angular resolution Mie scattering [15] calculations (0.18° cf. 1.0°) and a greater number of Monte Carlo simulations (10^8 cf. 10^5) than were possible for Ackleson and Spinrad [12], in order to generate an inversion lookup table (LUT) with 0.02 μm diameter and 0.001 RI resolution.
- (2) Development of an automatic clustering classifier to discriminate different phytoplankton types using flow cytometer data, specifically the separation between the globally significant cyanobacteria *Prochlorococcus* [16] and *Synechococcus* [17,18].

(3) Use of natural assemblages of phytoplankton (rather than cultures), sampled over 100° latitude through the Atlantic Ocean on the Atlantic Meridional Transect (AMT) research expeditions [19,20], spanning a decade (2008–2017), using a consistent instrument, operator, calibration, and standard operating procedure. This large dataset comprises order 10⁴ samples containing order 10¹⁰ individual cells.

In turn, these advances should have an impact on our understanding of likely cellular differences between different picoplankton species (size, size distribution, RI and, by inference, biochemical composition) as a function of depth and latitude.

2. METHOD

Two Becton Dickinson FACSort flow cytometers (serial numbers B0264 and B0043) with identical optical geometries were used to enumerate depth-resolved vertical profiles of picophytoplankton throughout the AMT time series. The individual samples were collected from Niskin bottles attached to an oceanographic rosette sampler into clean 250 mL polycarbonate bottles (Nalge Company, USA) at all available depths between the surface and 200 m. These were then stored at 4°C in the dark and analyzed within 2 h. Samples were analyzed at flow rates calibrated daily using Beckman Coulter Flowset fluorospheres of known concentration, prior to the analysis of seawater samples.

The flow cytometer used an air-cooled argon-ion laser (488 nm) in order to measure forward scatter ($\psi_{S3} = \pm 1.5 - 12.2^\circ$) and side scatter ($\psi_{S4} = 64 - 116^\circ$) (see schematic Fig. 1) red (>650 nm) and orange (585 ± 21 nm) fluorescence. Measurements of light scatter and fluorescence were pre-processed using CellQuest V3.3 software (Becton Dickinson, Oxford) with log amplification on a four-decade scale with 1024-channel resolution. Data were initially stored in Listmode format following Flow Cytometry Standard 2.0 and converted into ASCII using FlowPy (<http://flowpy.wikidot.com>) for onward analysis. The data extracted from the Listmode format, necessary for onward processing of the data, included the instrument gain settings, voltage settings for the fluorescence and scattering channels,

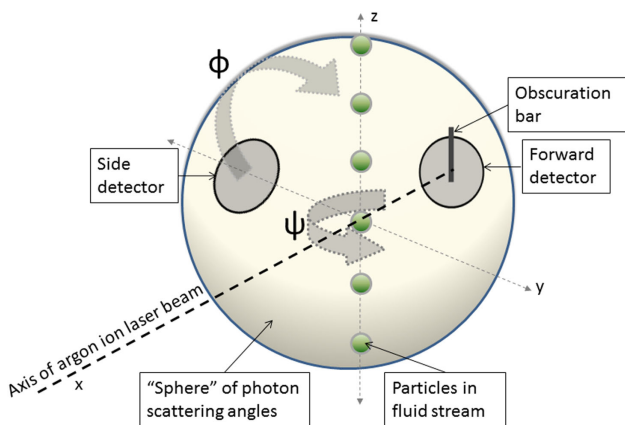


Fig. 1. Schematic diagram of scattering (ψ, ϕ) within the flow cytometer.

Table 1. Typical Flow Cytometer Settings Optimized for the Detection of Prokaryotic (Pico) and Eukaryotic (Nano) Algae During AMT^a

Detector	Pico Setting	Nano Setting
Forward scatter (S3)—Gain	1	10 ⁻¹
Side scatter (S4)—PMT voltage	440 V	265 V
Green fluorescence (FL1)—PMT voltage	400 V	400 V
Orange fluorescence (FL2)—PMT voltage	600 V	365 V
Red fluorescence (FL3)—PMT voltage	570 V	360 V

^a The Flowset sphere calibrations were carried out using the nano setting. The forward scatter setting is manifested as a gross gain (more sensitive for pico than nano).

flow rate, and length of analysis time (in order to work out the volume sampled) for each sample run.

Two instrument settings were used on the flow cytometer in order to discriminate different types of phytoplankton within the sample, these being labelled as nano and pico for eukaryotic and prokaryotic phytoplankton, respectively, in Table 1. The flow rate calibration spheres were analyzed using the nano settings.

A. Theory

The key premise of this paper is that it is possible to determine the size and RI of any particle from its measured scattering characteristics using a flow cytometer if it can be referenced against theoretically derived scattering values (Mie [15]) for a particle of known size and RI, for example, a polystyrene sphere.

Following the description and nomenclature of Ackleson and Spinrad [12] for a detector i and the intensity of scattered light from an unknown particle p , recorded by the flow cytometer, may be expressed as

$$S(i)_p = X(i)I(i)_p G(i)_p (dC_{sca}/d\Omega_i)_p, \quad (1)$$

where $S(i)$ is the channel value, $I(i)$ is the laser intensity, $G(i)$ is the detector gain factor, and $(dC_{sca}/d\Omega)$ is the differential scattering cross section. Here $X(i)$ may be considered a repository for all unknown factors such as filter transmissivity and lens efficiency.

For a reference particle r , the equivalent to Eq. (1) may be written as

$$S(i)_r = X(i)I(i)_r G(i)_r (dC_{sca}/d\Omega_i)_r. \quad (2)$$

Combining Eqs. (1) and (2) results in

$$(dC_{sca}/d\Omega_i)_p = \frac{S(i)_p I(i)_r G(i)_r}{S(i)_r I(i)_p G(i)_p} (dC_{sca}/d\Omega_i)_r. \quad (3)$$

The flow cytometer used in this study had a detector in the forward scatter direction [$S(3)$] and a detector in the side scatter direction [$S(4)$], where the nomenclature of $S(3)$ and $S(4)$ is adopted from Bohren and Huffman [15]. This simultaneously allowed the derivation of a differential scattering cross-section pair, $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$, given a

well-characterized instrument with frequently available reference particle measurements $[S(3)_r, S(4)_r]$, together with reference particle values $[(dC_{sca}/d\Omega_3)_r, (dC_{sca}/d\Omega_4)_r]$ calculated using the Mie scattering formalism. For further scattering calculation details, such as the relationship between the differential scattering cross section and particle phase function, determination of the angular weighting function specific to an individual flow cytometer using Monte Carlo modelling, and the calculation of a resultant Mie-theory-derived LUT, see Appendices A, B, and C respectively.

The theoretical calculations, specific to the wavelength and internal geometry of the flow cytometers used in this study, are presented in Fig. 2; this shows that for a range of particle sizes (approximately between 1 and 10 μm) and refractive indices (approximately between 1.01 and 1.09), that a unique solution exists for the particle size, RI pair for a given pair of $(dC_{sca}/d\Omega_3)_p$ and $(dC_{sca}/d\Omega_4)_p$. Outside of these ranges, harmonics in the Mie solution lead either to multiple solutions or small changes in $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$, which result in relatively large changes in size and RI.

The LUT was used in inverse mode, by minimizing the distance between the individually derived $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$ pair and theoretically calculated values, to determine particle size and RI. The implementation of Eq. (3) requires directly measured values of $S(3)_p$ and $S(4)_p$ and their associated gain settings of $G(3)_p$ and $G(4)_p$ as well as the nearest (in time, typically daily) flow-rate calibration run for beads of known diameter (3.6 μm) and RI (polystyrene RI = 1.198

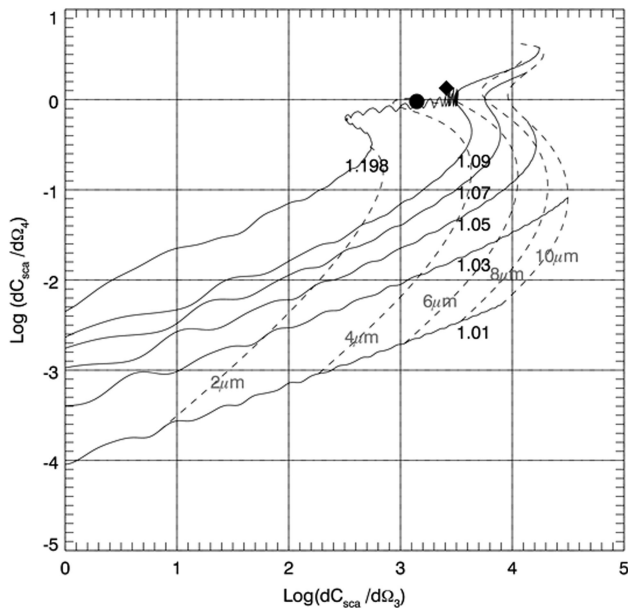


Fig. 2. Mie-theory-derived lookup table results of log-transformed differential side scatter $(dC_{sca}/d\Omega_4)$ against differential forward scatter $(dC_{sca}/d\Omega_3)$ for the flow cytometers used in this study. Mie scattering calculations were carried out at 0.180° resolution (number of angles = 999) for particles of diameter (D) ranging between 0.02–10 μm ($\Delta D = 0.02 \mu\text{m}$) and real RI between 1.010 and 1.200 ($\Delta \text{RI} = 0.001$). Solid contours are for constant RI from 1.01 to 1.09 in 0.02 increments; dashed contours are constant particle diameter from 2 to 10 μm in 2 μm intervals. Filled circles and diamonds are the theoretical results for a polystyrene bead (RI = 1.198) with a diameter of 3.6 and 5.16 μm , respectively.

cf. seawater) to determine $S(3)_r$ and $S(4)_r$, together with their associated gain settings $G(3)_r$ and $G(4)_r$.

However, upon implementation using the different settings of the flow cytometer (Table 1), it was found that the following additional considerations needed to be made:

- (1) The reference particle calibration was carried out using the nano settings (see Table 1), stressing the importance of using the gain ratio of $G(3)_r/G(3)_p$ in Eq. (3) in order to determine the forward scatter component $(dC_{sca}/d\Omega_3)_p$.
- (2) The side detector gains to determine $(dC_{sca}/d\Omega_4)_p$ were set to unity for both the pico and nano settings. However, the photomultiplier tube (PMT) voltage values for the side detector were generally different (Table 1) for these two settings. In order to accurately reference to $(dC_{sca}/d\Omega_4)_r$ for the pico settings, an effective gain (G_{eff}) needed to be calculated. This was calculated using $G_{\text{eff}} = 10^{(V_p/V_r)}$ where V_p and V_r are the PMT voltages for the particle and reference flow beads, respectively.

B. Cluster Analysis

The high-throughput nature of FC data carries the requirement that some form of cluster analysis is necessary in order to classify and enumerate particles. This can be subjective, based on expert manual interpretation using the imposition of gates and thresholds [10], or objective [21], where prior decisions need to be made concerning what constitutes a cluster (number of points, separation between points). In this paper we use an objective density-based approach for both the reference calibration and the separation of the different types of phytoplankton.

Four (1–4) grouping types were automated in this work: (1) Separating out the *Synechococcus* is achieved by using the pico settings (Table 1) for orange fluorescence (FL2) and $(dC_{sca}/d\Omega_4)_p$; and (2) by retaining the pico settings and using red fluorescence (FL3) against $(dC_{sca}/d\Omega_4)_p$ to discriminate *Prochlorococcus* plus *Synechococcus*. The contribution from *Prochlorococcus* alone is determined by subtracting the classifier (1) from (2). (3) Separating out the *Synechococcus* plus *Cryptophytes* is achieved by using the nano settings (Table 1) of the FC for FL2 and FL3. (4) Retaining the nano settings and using FL3 against $(dC_{sca}/d\Omega_4)_p$ allows everything else to be distinguished.

For each grouping type (1–4), the points that constitute the cluster are determined as follows: (i) the observations are re-binned onto an $x - y$ grid of dimensions 256×256 and a density field created using a 2D histogram function; (ii) the resulting field is sorted into descending values of density, i.e., how many data points fall within a grid cell, and a threshold (ϵ) set as

$$\epsilon = \sqrt{dx^2 + dy^2}, \quad (4)$$

where dx is the grid spacing in the x direction and dy in the y direction; (iii) the DBSCAN [22] clustering algorithm is applied to the data, and a cluster is defined as where there are more than 10 points in a grid cell, and its “connectivity” is defined by Eq. (4) (i.e., it is within the search radius of a single grid point from at least one other filled grid cell). Additionally, for grouping 1 (*Synechococcus*) a

threshold in FL2 was set to exclude values <1.0 , and for grouping 2 (*Prochlorococcus*) thresholds were set as follows: for a higher setting, selected by the original operator of the flow cytometer, $-3.0 \leq \log_{10}(dC_{sca}/d\Omega_4)_p \leq 2.0$ and $0.65 \leq FL3 \leq 2.5$; for a lower setting, where the operator is specifically enumerating low chlorophyll concentrations of *Prochlorococcus* $-3.5 \leq \log_{10}(dC_{sca}/d\Omega_4)_p \leq -0.5$ and $(FL3_{min} + 0.2)FL3 \leq 4.0$, where $FL3_{min}$ is the minimum red fluorescence within an individual sample dataset.

The cluster-analysis-determined abundances ($N \text{ mL}^{-1}$) for *Synechococcus* and *Prochlorococcus* were compared with abundances determined using a manually intensive gated approach [10] as a quality control check on the efficacy of the clustering technique.

3. RESULTS

A. Size Verification Against Polystyrene Beads

The efficacy of the particle size retrieval was tested using polystyrene beads over a range of sizes and a fixed RI ($RI = 1.198$) sampled using the FACSORT (SN B0264) instrument. The bead sizes tested were 1.41, 2.08, 3.1, 5.16, 9.99, 10.0, 16.7, 20, and 37 μm suspended in 0.1 μm filtered Milli-Q water at typical concentrations of $0.47 - 8.0 \times 10^8$ beads mL^{-1} . Each size sample was run at a low flow rate ($\sim 20 \mu\text{L min}^{-1}$) until 20,000 scattering events had been counted (typically over a period of 200–300 s). The values of $(dC_{sca}/d\Omega_3)_p$ and $(dC_{sca}/d\Omega_4)_p$ were then determined using a cluster analysis (see Section 2.B), and the range of variability was set where the density at the edge of the cluster fell below half the peak value (akin to full width at half-maximum). The results of the intercomparison are shown in Fig. 3.

The different sized bead values of $(dC_{sca}/d\Omega_3)_p$ and $(dC_{sca}/d\Omega_4)_p$ have been referenced to the 5.16 μm bead measurements $[(dC_{sca}/d\Omega_3)_r, (dC_{sca}/d\Omega_4)_r]$ in conjunction

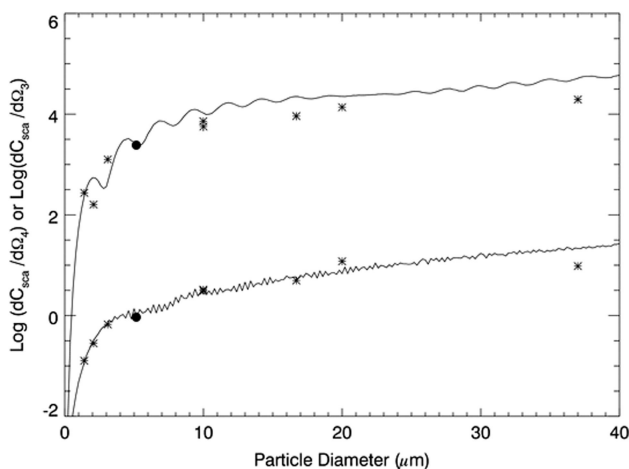


Fig. 3. Comparison of measured (data points) and predicted (solid lines) differential scattering cross section for polystyrene beads ($RI = 1.198$) of different sizes using a Becton Dickinson FACSORT flow cytometer. The upper curve is differential forward scatter $(dC_{sca}/d\Omega_3)_p$, and the lower curve is differential side scatter $(dC_{sca}/d\Omega_4)_p$. The filled circle represents the scattering data against which all the other bead measurements are referenced (5.16 μm diameter).

with Mie scattering theory to allow the derivation of a $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$ pair for each particle size. As is expected from the theory of particle scattering, the particle phase function is strongly peaked in the forward direction, resulting in the values of $(dC_{sca}/d\Omega_3)_p \gg (dC_{sca}/d\Omega_4)_p$ by around 3 orders of magnitude. Both curves show characteristic Mie scattering harmonics, this being particularly pronounced in $(dC_{sca}/d\Omega_3)_p$ at sizes below 10 μm . The intercomparison between $(dC_{sca}/d\Omega)_p$ and Mie scattering theory is closer for side scattering than forward scattering. The log-transformed root mean squared differences (log-RMSD) in size between theory and observation for all sizes (excluding the reference particle) were 0.158 and 0.363 for $(dC_{sca}/d\Omega_4)_p$ and $(dC_{sca}/d\Omega_3)_p$, respectively. When the size range is restricted to $\leq 10 \mu\text{m}$ (i.e., the equivalent size of phytoplankton cells relevant to this study), then there is a lower log-RMSD for $(dC_{sca}/d\Omega_4)_p$ of 0.047, whereas there is a marginal worsening for $(dC_{sca}/d\Omega_3)_p$ to 0.369. These differences are not reconcilable by varying the RI of the particles or by inspecting the statistical variability in the scatter for the beads (indeed, the observational variability, although plotted in Fig. 3, is almost invisible). It is particularly challenging to directly compare the model predicted against actual bead diameter because of the relatively high RI of the reference particle. For small ($<2 \mu\text{m}$) particle sizes, $(dC_{sca}/d\Omega_4)_p / (dC_{sca}/d\Omega_3)_p$ tends towards the same value over a wide range of refractive indices (see Fig. 3 of Ackleson and Spinrad [12] and Fig. 2); for higher refractive indices, multiple superimposed harmonics for small changes in particle physical characteristics make determining size by inverting the forward and side scatter measurements challenging and often ambiguous. For the limited number of data point pairs shown here (Fig. 3), and under the challenging conditions of a relatively high RI, the root mean square error (RMSE) in retrieving particle size ($<5 \mu\text{m}$) from a $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$ pair is $\pm 0.33 \mu\text{m}$.

However, interpreting both Figs. 2 and 3 together shows that: (i) it is possible to use a calibration run of polystyrene beads to act as a reference point (filled circle in Fig. 2) and (ii) over the range of interest for marine picoplankton size ($\leq 10 \mu\text{m}$) and RI ($RI = 1.01 - 1.09$) unique solutions are possible from direct measurement of forward and side scatter.

B. Refractive Index Verification Against Suspended Oils

Emulsion samples of different hydrocarbons (Table 2) were prepared by adding 150 μL of oil to 1.35 mL of artificial seawater (ASW, 34 ‰ salt content) and shaking vigorously to produce droplets over a wide size range continuum. Emulsions were analyzed in the FC using the same sheath fluid as oil suspension medium (i.e., ASW). Initial observation showed that the shorter-chain hydrocarbons tended to de-emulsify quickly. Therefore, the samples were paused and re-agitated after 50,000 events were counted in order to refresh the emulsion.

Figure 4 shows that the observations in general are parallel to the theoretical curves of constant RI, with a clear distinction between the different hydrocarbon emulsions in terms of RI. The average and standard deviation over the LUT-retrieved RI was constrained between diameters of ≥ 1 and $\leq 4 \mu\text{m}$ and

Table 2. Comparison Between Literature Values of RI for Different Oil Types and Those Derived from the Flow Cytometer Observations^a

Oil Type	RI	RI Relative to Seawater	FC Derived RI ($\pm\sigma$)	Number of Observations
Hexane	1.3806 ^a	1.0272	1.0322(\pm 0.0188)	14793
Heptane	1.3934 ^b	1.0367	1.0357(\pm 0.0136)	15131
Octane	1.4035 ^b	1.0443	1.0384(\pm 0.0089)	14919
Nonane	1.4058 ^c	1.0460	1.0416(\pm 0.0091)	21411
Decane	1.4102 ^c	1.0492	1.0439(\pm 0.0057)	16549
Dodecane	1.4210 ^c	1.0573	1.0485(\pm 0.0065)	23169
Tetradecane	1.4290 ^c	1.0632	1.0515(\pm 0.0066)	16712
Pentadecane	1.4315 ^c	1.0651	1.0528(\pm 0.0089)	15821

^a [a] $\lambda = 488$ nm, $T = 20^\circ\text{C}$ [23]; [b] $\lambda = 488$ nm, $T = 20^\circ\text{C}$ [24]; [c] $\lambda = 589$ nm, $T = 20^\circ\text{C}$ [25]. RI is given relative to the refractive index of seawater (1.344 Austin and Halikas [26]).

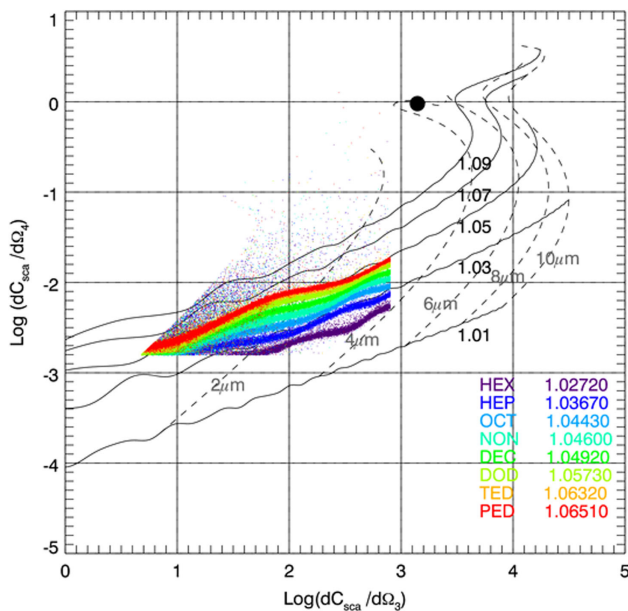


Fig. 4. Flow cytometric measurements of differential scattering cross section from eight different oil emulsions suspended in artificial seawater (see Table 2 for further details). HEX, hexane; HEP, heptane; OCT, octane; NON, nonane; DEC, decane; DOD, dodecane; TED, tetradecane; PED, pentadecane. The filled circle is the reference measurement for a bead of diameter $4\ \mu\text{m}$ and RI 1.198.

truncated where obvious Mie harmonics were visible in the data ($dC_{\text{sca}}/d\Omega_3 \leq 800$ in Fig. 4). The range of precision in the retrievals was between ± 0.0057 (decane) and ± 0.0188 (hexane), although the poorer results for hexane are likely due to the lower $dC_{\text{sca}}/d\Omega_4$ signal and the tendency for the shorter-chain hydrocarbons to quickly de-emulsify. The average precision over the range of RI is of order ± 0.01 . The range in accuracy was between ± 0.001 (heptane) and ± 0.0123 (pentadecane) with an average of ± 0.0068 , with a tendency for the longer-chain hydrocarbons to have a greater discrepancy from theory. Some of these tendencies may be explained by variations in temperature during the period of analysis within the FC, which would reduce RI with increasing temperature ($T \geq 20^\circ\text{C}$). However, there is an underestimate of RI in Table 2 for nonane to pentadecane, as the only literature values [25] to our knowledge are reported at 589 nm (cf. laser $\lambda = 488$ nm), so the wavelength and temperature effects may counteract each other.

C. Size and Refractive Index Retrievals of Natural Picoplankton Assemblages

Figure 5 shows the efficacy of the clustering algorithm for both a) *Synechococcus* and b) *Prochlorococcus*. For both species, the clustering algorithm is on the conservative side, tending to underestimate the concentration compared with the manual gated approach. For the specific example here, *Synechococcus* concentrations are 14,458 cells mL^{-1} and 21,069 cells mL^{-1} for the clustering and gated methods, respectively ($\sim 69\%$ agreement); for *Prochlorococcus* the values are 61,587 and 70,787 cells mL^{-1} , respectively ($\sim 87\%$ agreement). However, most of the cells omitted from the automated approach are on the fringes of the cluster, so it is likely that this does not greatly impact the enumeration of the size distribution shape and associated RI. For the entire AMT FC dataset analyzed in this paper (10 transects, 7413 samples), the clustering and gated concentrations were compared for both *Synechococcus* and *Prochlorococcus*. For data where the clustering and gated concentrations were within a factor of 10, a log-log regression analysis resulted in: *Synechococcus* slope = 0.824, intercept = 0.307, $R^2 = 0.759$, $N = 5699$; *Prochlorococcus* slope = 0.941, intercept = 0.029, $R^2 = 0.901$, $N = 5768$. The corresponding density plots for size and RI, plotted in log $[(dC_{\text{sca}}/d\Omega_3)_p, (dC_{\text{sca}}/d\Omega_4)_p]$ space, are shown in Fig. 5(c) for *Synechococcus* and Fig. 5(d) for *Prochlorococcus*.

For *Synechococcus* there is a clear single peak in the size distribution around $1.75 \pm 0.25\ \mu\text{m}$, whereas for *Prochlorococcus* there is a double peak: one at $0.75 \pm 0.25\ \mu\text{m}$, where the concentration of particles is greatest (1.7×10^4 cells mL^{-1}), and another at $2.25 \pm 0.25\ \mu\text{m}$ (1.2×10^4 cells mL^{-1}).

This is shown more clearly in Fig. 6 where, additionally, the corresponding distribution of RI is plotted. For *Synechococcus* the peak in the RI is at 1.04, whereas the highest peak in RI for *Prochlorococcus* is at 1.09 (1.7×10^4 cells mL^{-1}) with the secondary lower peak at 1.03 (1.2×10^4 cells mL^{-1}). The *Prochlorococcus* double peak is present in the entire AMT dataset, implying two subpopulations or possibly ecotypes [27–29] throughout the water column that physically differ in terms of their size and optical properties (and, by inference, their biochemical composition [30]). The double peak is variable with depth and latitude as outlined in Section 3.D below.

The phytoplankton enumerated using the nano settings (Table 1) on the flow cytometer are shown in Fig. 7 for the same sample described in Fig. 5. Two clear populations are visible

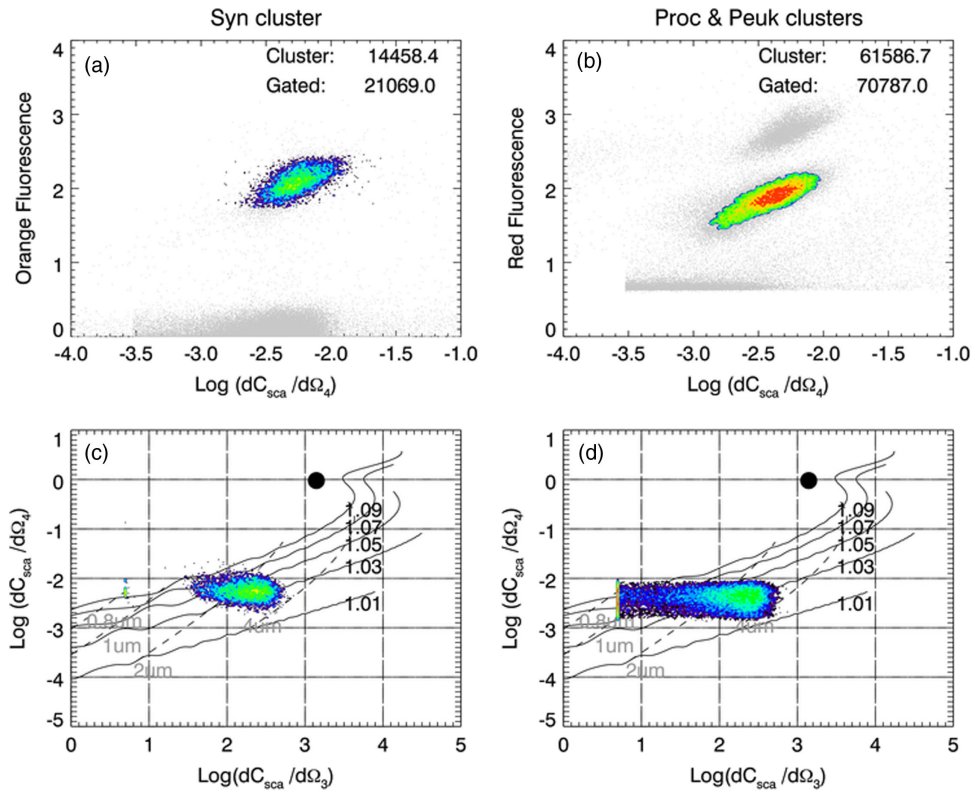


Fig. 5. Cluster analyses using the pico settings (Table 1) for (a) *Synechococcus* (Syn) and (b) *Prochlorococcus* (Proc). Grey symbols indicate all data points, and colored density plots show cluster membership, where warmer colors indicate a greater number of particles. Resultant pairs of $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$ as a density plot for (c) *Synechococcus* and (d) *Prochlorococcus*, superimposed on contours of size and RI. The filled circle represents the reference particle. Sample taken on AMT18 at 12:42 GMT on 6 October 2008 at a depth of 10 m, location 49.37°N, 11.39°W.

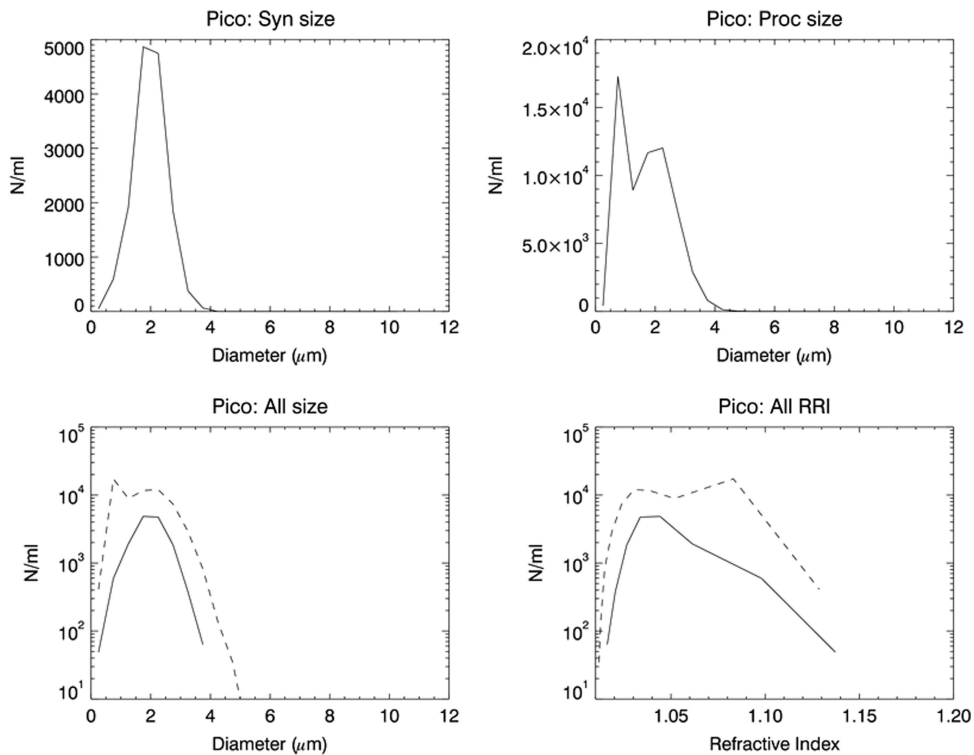


Fig. 6. Top panels: size distribution for *Synechococcus* (Syn) and *Prochlorococcus* (Proc) for the same sample shown in Fig. 5. Bottom panels: a direct intercomparison of (left) size and (right) RI for *Synechococcus* (solid line) and *Prochlorococcus* (dashed line). Note the log scales on the y axis.

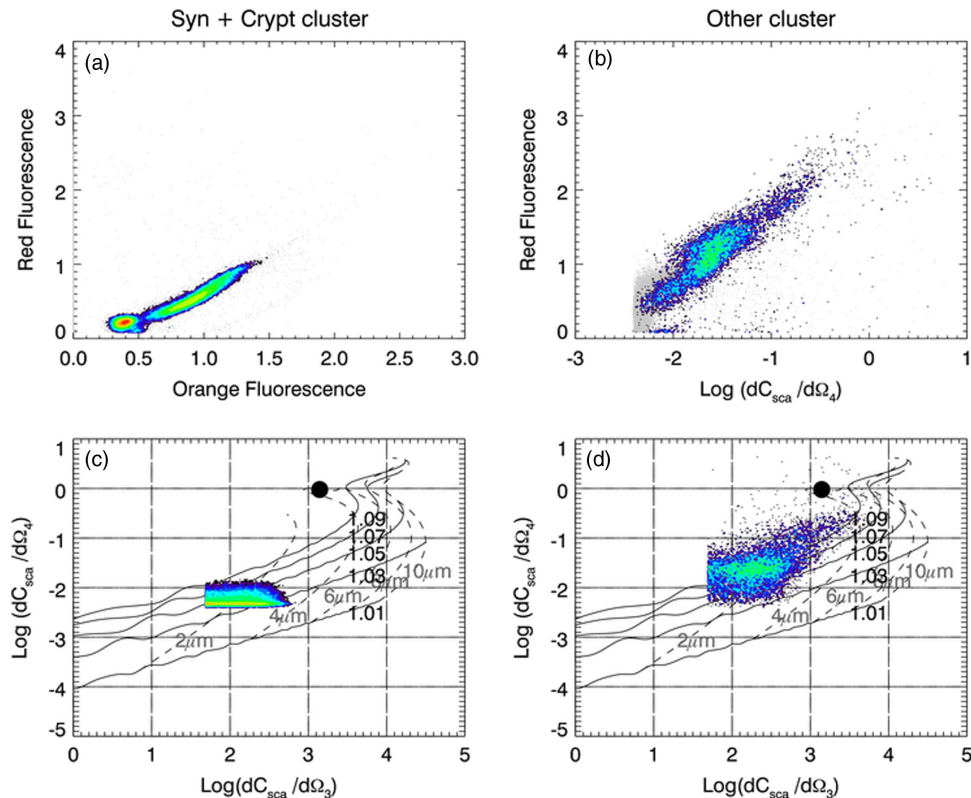


Fig. 7. Cluster analyses using the nano settings (Table 1) for (a) *Synechococcus* (Syn) plus *Cryptophytes* (Crypt) and (b) Other, otherwise unclassified phytoplankton. Grey symbols indicate all data points, and colored density plots show cluster membership, where lighter colors indicate a greater number of particles. Resultant pairs of $[(dC_{sca}/d\Omega_3)_p, (dC_{sca}/d\Omega_4)_p]$ as a density plot for (c) *Synechococcus* (Syn) plus *Cryptophytes* (Crypt) and (d) Other superimposed on contours of size and RI. The filled circle represents the reference particle. Sample taken on AMT18 at 12:42 GMT on 6 October 2008 at a depth of 10 m, location 49.37°N, 11.39°W.

in Fig. 7(a), with the *Synechococcus* cluster concentrated in the region approximately bounded by $FL2 < 0.5$ and $FL3 < 0.3$ and *Cryptophytes* in a more elongated section above this. The peak of the size distribution is around $1.75 \pm 0.25 \mu\text{m}$ with a corresponding RI of 1.04. Removing these cells from the analysis results in all “other” (and therefore unclassified) phytoplankton. There are likely multiple populations of different phytoplankton types in this part of the sample. Although the peak in the size distribution is around $1.25 \pm 0.25 \mu\text{m}$, there are multiple peaks in the RI between 1.07 and 1.15, suggesting contrasting biochemical compositions of the cells under analysis. The multiple groupings in this region of Fig. 7(b) are difficult to cluster into well-defined classifications because they do not have a signature fluorescence and/or size characteristic. The cellular concentrations of any grouping will also be a factor in this.

D. Atlantic Basin Wide Retrievals of Picoplankton Size and Refractive Index

The variability of *Prochlorococcus* abundances along the AMT (Fig. 8) is consistent with previous findings [10,19], namely, that *Prochlorococcus* reaches its highest abundance in the equatorial convergence between 20°N to the equator, with consistently high abundances in the nutrient-poor [31] surface layers of the North (40°N–12°N) and South (6°S–40°S) Atlantic Gyres

[32]. Using the scheme developed in this work, we can further show the size and RI characteristics of *Prochlorococcus* as a function of both latitude and depth. The peak diameter and RI were calculated as follows: (i) a histogram at 0.50 μm binned resolution (e.g., Fig. 6) was calculated for each type within an individual sample and the peak in the abundance identified; (ii) within the abundance peak bin, the mean (and associated standard deviation) diameter was calculated, and this value was reported as the peak diameter; (iii) the RI values associated with the peak abundance bin were used to calculate the mean (and associated standard deviation) RI, and this value was reported as the peak RI. The mean diameter and RI were calculated from the entire distribution, weighted by the abundances in each histogram size bin.

The smallest *Prochlorococcus* cells in Fig. 8 are concentrated just beneath (75–150 m) the maximum abundances of the equatorial upwelling region (10°N); here the peak diameters are $\leq 0.65 \mu\text{m}$. Slightly larger cells ($\sim 0.95 \mu\text{m}$) are situated in the centers of the Northern and Southern Gyres at a depth of ~ 50 m. Where the peak (or mode) significantly differs ($\geq 1.25 \mu\text{m}$) from the mean of the size distribution, this may be interpreted as either the presence of two size peaks, typical in our observations for *Prochlorococcus* (see Figs. 5 and 6) or a long tail in the distribution.

Figure 9 confirms the presence of two peaks ($\sim 0.75 \mu\text{m}$, $\geq 1.75 \mu\text{m}$) in the size distribution across the whole AMT,

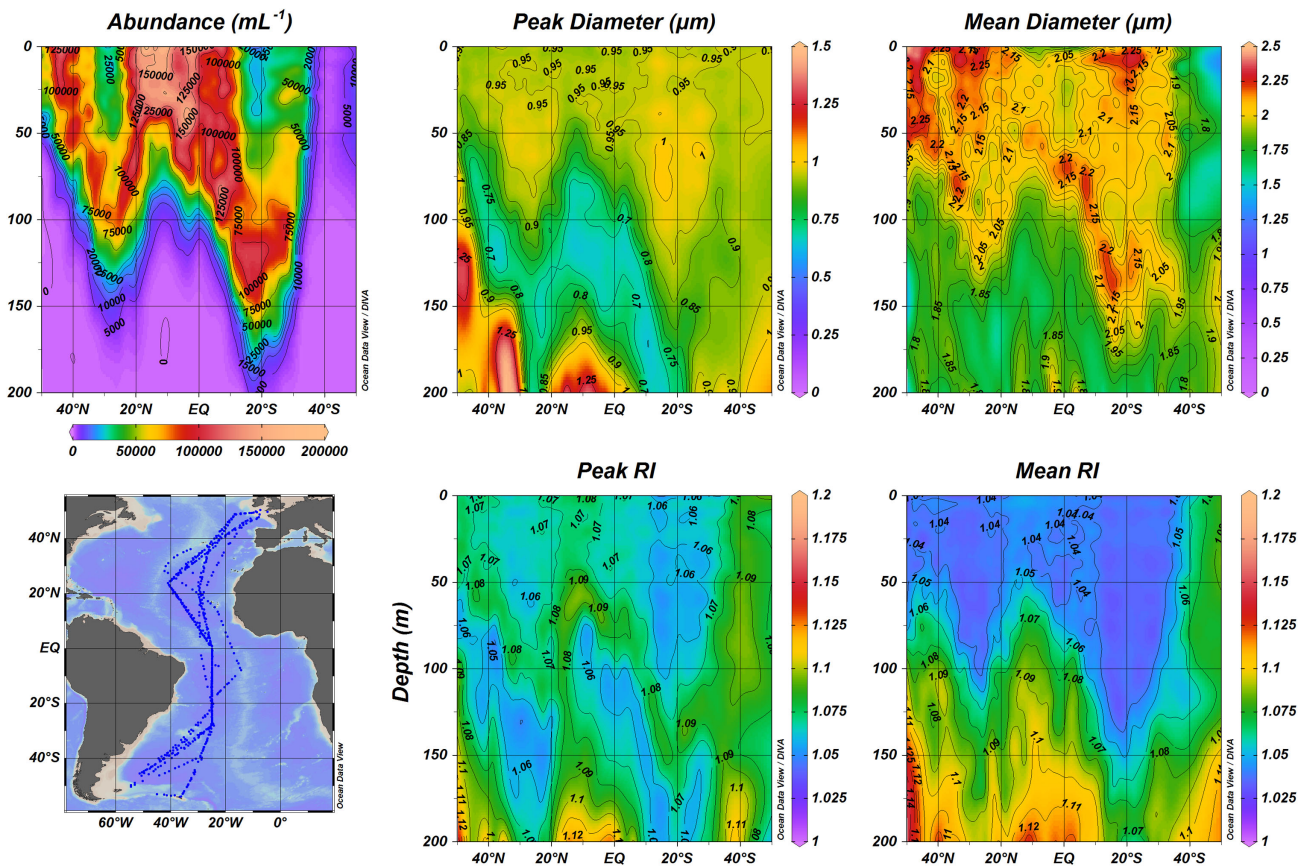


Fig. 8. Estimated abundance, size, and RI characteristics of *Prochlorococcus* as a function of depth and latitude (N to S) for AMT cruises 18–27 (2008–2017).

most pronounced in the surface 100 m (in general). This is shown most clearly in the South Atlantic Gyre (SATL) province [32], where the $0.75 \mu\text{m}$ peak dominates the $1.75 \mu\text{m}$ peak; the height of the $1.75 \mu\text{m}$ peak is between 40–50% of the $0.75 \mu\text{m}$ peak for the surface 50 m. As depth increases, the relative importance of the $1.75 \mu\text{m}$ peak increases to 60%–80% of the $0.75 \mu\text{m}$ peak as well as larger sizes ($>3 \mu\text{m}$), albeit in vanishingly small concentrations, becoming more common. A tendency towards monodispersion with depth is also shown: the $200(\pm 20)$ m depth-averaged size distribution for the SATL province shows a single peak at $1.25 \mu\text{m}$. Interestingly, the South Sub-tropical Convergence (SSTC) province gives a large (1.75 – $2.25 \mu\text{m}$) single peak in the surface 50 m, with the smaller peak (at very low concentrations) being confined to depths >100 m.

The surface values of RI (Fig. 8) for the peak and mean of the distribution are ~ 1.06 and 1.04 , respectively, for large parts of the AMT, with local minima in the centers of the gyres. In general there is an increase in RI with depth, with the highest values encountered at depths >150 m being ~ 1.12 .

The highest abundances of *Synechococcus* ($>15,000 \text{ cells mL}^{-1}$) are situated to the south of the South Atlantic Gyre (Fig. 10). It is here that the largest cells are encountered (~ 2.25 – $2.50 \mu\text{m}$). Conversely, the lowest surface concentrations of *Synechococcus* are where the smallest cells are encountered ($\sim 1.5 \mu\text{m}$) in the northernmost parts of the South

Atlantic Gyre. Unlike *Prochlorococcus*, there is little difference between the peak and mean of the size distribution; we can therefore conclude that *Synechococcus* size distributions are generally strongly monodispersed. There also seems to be more variability in the RI, with surface values (top 50 m) in the regions of highest concentration being ~ 1.04 , and those in the regions of lowest concentration being ~ 1.08 . Abundances decrease with increasing depth, and RI increases with depth (>150 m; $\text{RI} \sim 1.12$).

The unclassified picoplankton abundance, size, and RI characteristics (Fig. 11) appear to be strongly linked to the likely location of the deep chlorophyll maximum (DCM). In the Northern Gyre, the DCM is located around 125 m; the DCM then shoals in the region of the equatorial upwelling to just beneath 50 m (here the greatest abundances of $>20,000 \text{ cells mL}^{-1}$ are encountered); in the Southern Gyre the DCM reaches its greatest depths of ~ 150 m in the center of the gyre. In the mid-latitudes, both north and south, it is questionable that the DCM exists as a discernable entity, with the highest abundances ($\sim 10,000 \text{ cells mL}^{-1}$) are confined to the surface 50 m. The largest cells in Fig. 11 are encountered in the equatorial upwelling region ($>2.5 \mu\text{m}$); it is here that the size distribution is most monodispersed for this class of phytoplankton. In contrast, the region just above the DCM, particularly in the area between the equator and the center of the

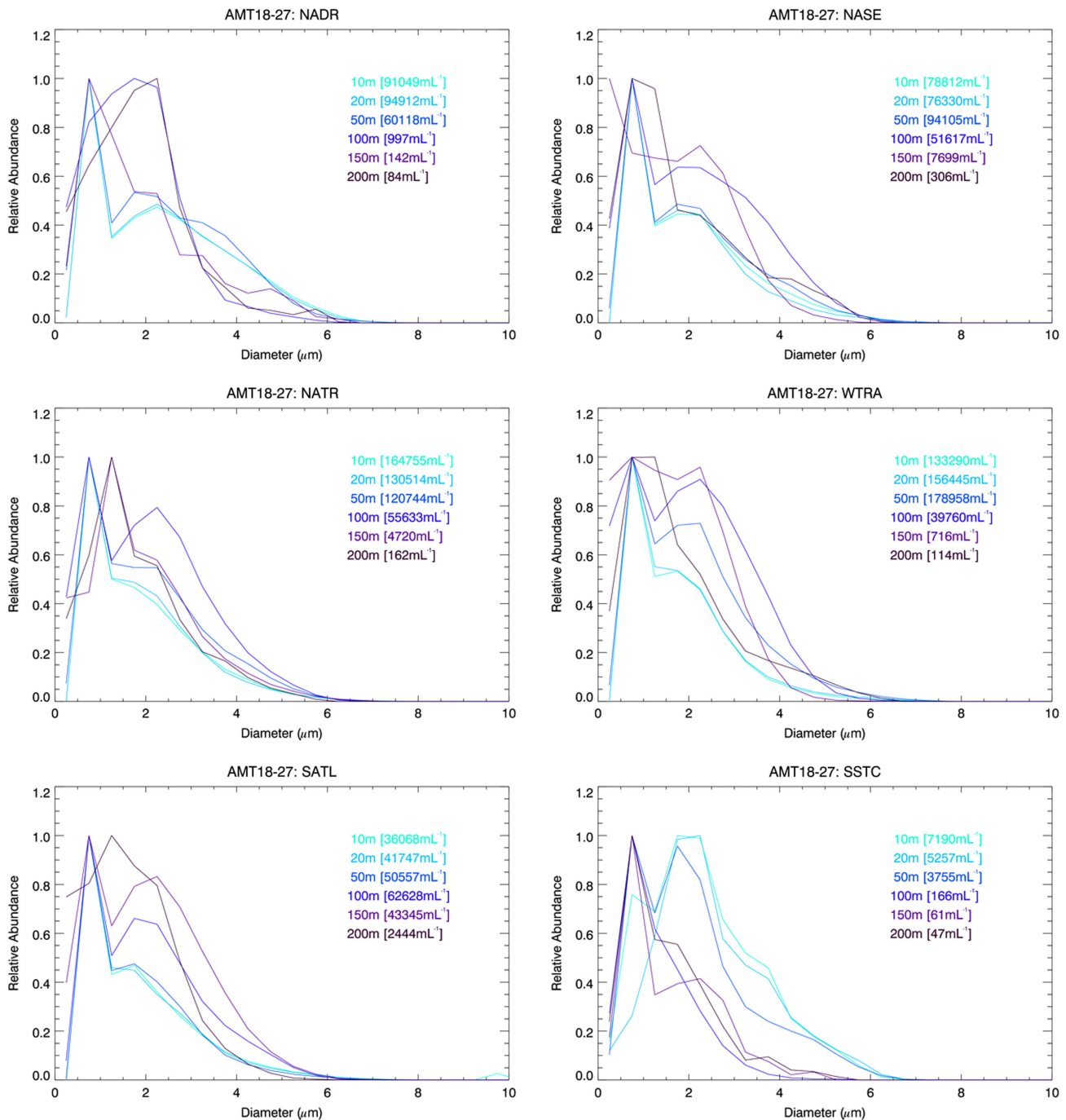


Fig. 9. Average depth-resolved size distributions at $10(\pm 5)$ m, $20(\pm 5)$ m, $50(\pm 10)$ m, $100(\pm 10)$ m, $150(\pm 20)$ m, and $200(\pm 20)$ m for six different provinces [32] of the Atlantic Ocean traversed by the AMT [19]. These are, from north to south: (i) North Atlantic Drift Region (NADR); (ii) North Atlantic Sub-tropical Gyre, East (NASE); (iii) North Atlantic Tropical Gyre (NATR); (iv) Western Tropical Atlantic (WTRA); (v) South Atlantic Gyre (SATL) and; (vi) South Sub-tropical Convergence (SSTC). For a map of the different provinces, see Fig 1. of Smyth *et al.*, (2017)[19]. Numbers in the square brackets are the average total concentrations of *Prochlorococcus* (mL^{-1}) over each depth interval.

Southern Gyre, shows the greatest offset ($1.25 \mu\text{m}$) between the peak and mean of the size distribution, which suggests multiple populations of possibly different species of picoplankton. The minimum in RI (< 1.05) closely follows the position of the DCM in the water column. The surface values are generally > 1.09 in both the peak and mean of the associated size distribution.

4. DISCUSSION

It has been customary [33] to calculate phytoplankton cell size from the forward light scattering signal from a flow cytometer, assuming a RI for the cell relative to water. In this paper, we have used forward and side scatter signals to infer both RI and particle size. The method is based on theoretical calculations of Mie scattering, according to which unique solutions exist for size and

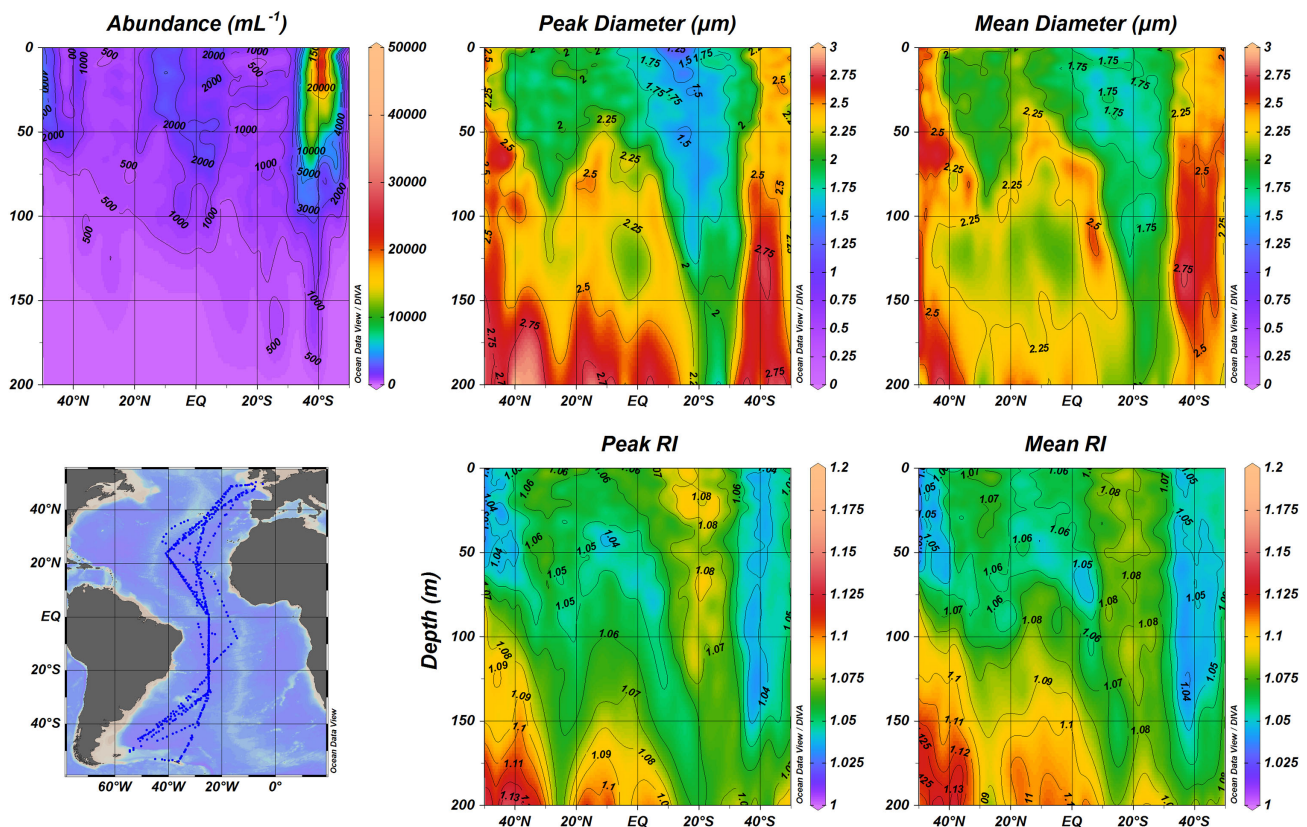


Fig. 10. Estimated abundance, size, and RI characteristics of *Synechococcus* as a function of depth and latitude (N to S) for AMT cruises 18–27 (2008–2017).

RI of particles in the relevant ranges, and tests of the theoretical results using reference materials have demonstrated the validity of the approach for the problem at hand and established the uncertainties in the results.

The method applied to depth-resolved FC data collected over a decade along the AMT has uncovered the richness of the vertical structure in picoplankton peak size, size distribution, and RI across the Atlantic basin. The typical peak sizes for *Prochlorococcus* reported here vary between 0.65 (equatorial upwelling, ~120 m) and 0.95 μm (center of the Northern and Southern Atlantic Gyres), which are slightly larger than previously reported (0.5–0.7 μm in diameter and 1.0 μm in length) [34]. For the larger *Synechococcus* picoplankton, the sizes vary between 1.50 (surface Southern Atlantic Gyre) and 2.75 μm (at depths > 150 m). Again, these are larger than previously reported in the literature, where they have been variously described as coccoid- to rod-shaped, 0.7 to 0.9 μm in diameter and 1.25–2.5 μm in length [35].

These differences from the earlier publications can be reduced if we were to assume lower RI values than those inferred directly using the method presented here. These discrepancies may also result, at least partially, from our assumption of homogeneous spherical cells (Mie theory restriction), as clearly both *Prochlorococcus* and *Synechococcus* are far from spherical [34,35]. The difference with the scattering calculations for an oblate spheroid are likely to be small (~2%) [36], but closer agreement can be found if the equivalent spherical diameter of the picoplankton cells is calculated. Algal cell structural differences

may cause a more significant deviation from theory [37,38]. Organelli, *et al.* [37] showed that backscatter predicted for homogeneous spheres is significantly less than that predicted for coated spheres. If this result were also applicable to side and forward scatter, a more realistically modeled algal particle, incorporating a cell wall with a higher RI compared with the cell interior, would yield smaller cell sizes than those predicted assuming a homogeneous sphere. One of the difficulties with implementing such a model is associated with calibration of the flow cytometer, which is typically carried out using homogeneous spheres (e.g., polystyrene bead) of known RI; to relate that to picoplankton cells with unknown interior structures poses perhaps too complex a problem for a quantitative solution. Interestingly, the RI used by Organelli, *et al.* [37] for the AMT26 cruise was 1.06. This is close to the values found here in the surface layers of the Atlantic Ocean for both *Prochlorococcus* (1.06–1.08; Fig. 8) and *Synechococcus* (1.04–1.08; Fig. 10).

Regardless of the absolute size of the cells, our results unambiguously show that clusters identified as *Prochlorococcus*, by both the conventional gating method and the automated clustering method, are frequently in fact a combination of two populations with distinct side scatter, forward scatter combinations (let us call them “opto-types”). The Mie scattering calculations suggest that one opto-type is characterized by higher RI and smaller size than the second opto-type. Their distributions are consistent with those of previously known ecotypes [27–29] of *Prochlorococcus* with distinct adaptations to environmental factors, such as light intensity, temperature, and

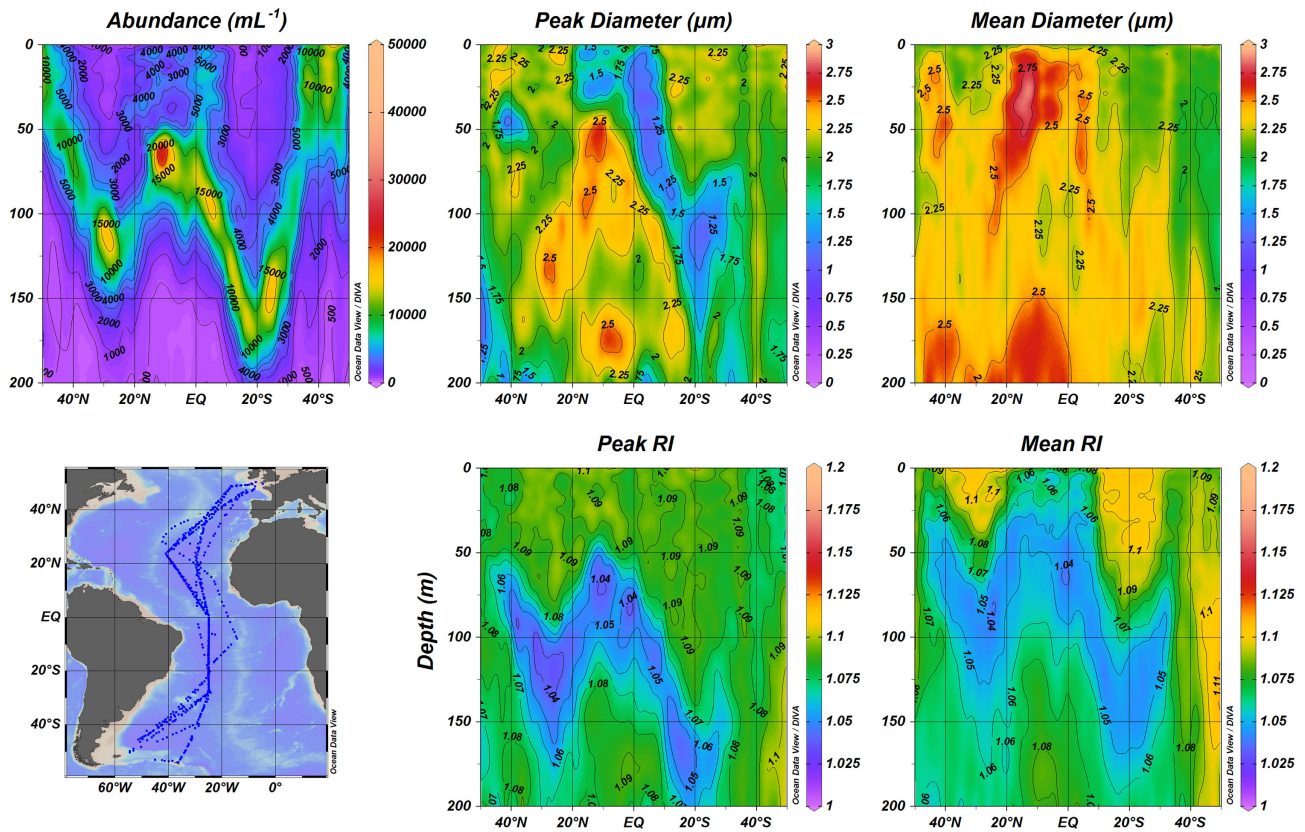


Fig. 11. Estimated abundance, size, and RI characteristics of unclassified picoplankton as a function of depth and latitude (N to S) for AMT cruises 18–27 (2008–2017).

nutrient concentrations [34], as well as remarkable genetic and physiological diversity [27]. The two opto-types are characterized by distinct Gaussian distributions in RI and size, often with overlapping tails, indicating that the two populations can coexist, with their relative abundance changing with environmental conditions. The method introduced here has allowed them to be mapped at basin scales for the first time, based uniquely on FC observations (Fig. 6), revealing the details of their distribution at the scale of the entire basin (Figs. 8 and 9), only made possible by having 10 years' worth of data. The differences between the mode and the mean values of RI and size indicate the importance of the population associated with the secondary peak, relative to the dominant population responsible for the mode.

That the values of RI are distinct for the two opto-types is particularly interesting, since values of RI may also hold clues to the biochemical makeup of the algal cells [30]. If changes in RI could indeed be related to the composition of organic material in the cells, then it offers the exciting possibility of interrogating the cells for some information on their chemical composition and its variation with latitude and depth using FC. The carbon and chlorophyll content of individual cells has previously been empirically related to the real and imaginary parts of the RI, respectively [39–41]. While these empirical relationships may only be applicable to open ocean [39] or more turbid [41] marine environments, generally an increase in the real part of the RI is associated with an increase in the intracellular carbon content. As the imaginary part of the RI has not been calculated in our approach, the determination of carbon-to-chlorophyll

ratios was not attempted. However, this may be an avenue of research in the future, possibly by calibrating the fluorescence channels of the flow cytometer to a chlorophyll concentration.

The stoichiometry of individual cells may be theoretically possible, as the RI of the individual building blocks of algal cells has been previously determined. Aas [30] gave the following refractive indices: opal (biogenic silica—bSi) ~ 1.07 ; lipid ~ 1.10 ; carbohydrate ~ 1.15 ; calcite ~ 1.16 ; and protein ~ 1.20 . In reality, algal cells are an unknown mixture of these components, which makes the calculation of a typical RI difficult using stoichiometry alone (see Laws [42] for a calculation of open-ocean photosynthetic quotients). Indeed, the internal cellular structure may be more important in determining RI than its internal composition, in particular the cell wall thickness and its RI relative to the internal components (lipid, carbohydrate, protein). The cell wall of *Synechococcus* is primarily protein [43], which would suggest a higher RI than is observed in the field (Fig. 10), leading to the obvious conclusion that composition and structure are both essential for determining cellular RI.

The wider inorganic nutrient stoichiometry of the Atlantic Ocean is also likely to play a role in determining which picoplankton dominate where. For example, the AMT dataset shows a deficit of nitrate to phosphate (i.e., a deviation from the Redfield ratio [44]) in the surface layers (0–50 m) in all of the Atlantic provinces between 40°N and 40°S [19]. This is particularly pronounced in the Southern Hemisphere, and in the South Atlantic Gyre the N:P ratio shows a nitrogen deficit down to depths >200 m [19]. It is this region that the smallest

picoplankton ($\sim 0.95 \mu\text{m}$), with the lowest RI (1.06), inhabit (Fig. 8). Much of the Atlantic exhibits a deficit in silicate [19], particularly at depths between 50–150 m; this partially explains why diatoms, whose cell walls are made of silica, are almost absent from the majority of the AMT transects. A simplistic analysis of RI by latitude and depth profiles would lead to the conclusion that where RI ~ 1.07 , the outer cell structure is dominated by biogenic silica, whereas values ~ 1.16 are likely dominated by calcite (e.g., coccolithophores).

5. CONCLUSIONS

The applicability of FC to determining the key properties of size and RI of natural picoplankton assemblages across the Atlantic basin has been demonstrated. The technique described has been applied to $\sim 10^4$ samples using a combination of an automated clustering analysis to differentiate between globally ubiquitous *Prochlorococcus* and *Synechococcus* with Mie scattering theory in order to obtain individual cell size and RI. The *Prochlorococcus* size distributions show distinct populations (double peak) in terms of size and refractive index, likely corresponding to different ecotypes dominating in different parts of the water column, whereas *Synechococcus* tends towards monodispersity. The peak sizes for *Prochlorococcus* range between 0.65–0.95 μm , which is larger than previously reported, and RI of between 1.06 and 1.12. Size, size distribution, and RI vary strongly with depth. *Synechococcus* are larger in size than *Prochlorococcus*; they range between 1.5–2.75 μm and have a RI of between 1.04 and 1.13.

APPENDIX A: SCATTERING THEORY

The volume scattering function (β) of a particle, scattering from direction ξ' into ξ for subsequent observation by a detector of solid view angle $d\Omega$, may be described as

$$\beta(\xi' \rightarrow \xi) d\Omega(\xi) = \beta(\psi, \phi) \sin \psi d\psi d\phi, \tag{A1}$$

where (ψ, ϕ) is the scattered direction of photons in a coordinate system centered on incident direction ξ' (Fig. 1). In natural waters, β depends only on the scattering angle ψ , and therefore Mie scattering calculations need only be carried out as a function of ψ as

$$\beta(\psi, \phi) = \beta(\psi), \tag{A2}$$

where ψ and ϕ are independent random variables that may be determined using a Monte Carlo simulation. As they are independent, Eq. (A1) may be written as a product of probability density functions (pdfs), and thus

$$\beta(\psi) \sin \psi d\psi d\phi \equiv p_\psi(\psi) d\psi \cdot p_\phi(\phi) d\phi. \tag{A3}$$

For the zenithal angle ϕ , assume, for a uniform distribution, that

$$p_\phi(\phi) d\phi = \frac{1}{2\pi} d\phi, \tag{A4}$$

and within a Monte Carlo simulation, the zenith angle can be determined by selecting a random number R_ϕ [0,1] as

$$\phi = 2\pi R_\phi. \tag{A5}$$

For the azimuth angle ψ , the pdf is defined as

$$p_\psi(\psi) = 2\pi\beta(\psi) \sin \psi. \tag{A6}$$

In order to determine the azimuthal angle using a Monte Carlo simulation, a random number R_ψ [0,1] is selected, and the integral is solved numerically to determine ψ corresponding to R_ψ as follows:

$$R_\psi = 2\pi \int_0^\psi \beta(\psi) \sin \psi d\psi, \tag{A7}$$

with the condition that

$$2\pi \int_0^\pi \beta(\psi) \sin \psi d\psi = 1, \tag{A8}$$

which is in the form of a cumulative distribution function.

In this paper, the Mie scattering code of Bohren and Huffman [15], written in the IDL programming language, was used to determine the particle phase (volume scattering) function. The Mie code was used in two ways. First, to calculate the weighting function for the forward and side detectors, as not all photons directed at azimuth angle ψ within the acceptance angles for each detector will necessarily end up being observed because of rotational symmetry about ϕ (see Fig. 1). Second, to create a LUT of forward and side differential scattering cross-section pairs ($dC_{\text{sca}}/d\Omega_{S3}$, $dC_{\text{sca}}/d\Omega_{S4}$), which can then be used (inverted) to determine an individual particle's size (D) and refractive index (RI).

APPENDIX B: CALCULATION OF THE WEIGHTING FUNCTION

The reason for determining the weighting function is to account for the fraction of photons that are scattered at azimuth angle ψ (Mie scattering in two dimensions) and in a direction ϕ (i.e., three dimensions) such that it will impact the forward or side detectors. A single Mie scattering calculation was carried out for a particle of $D = 4 \mu\text{m}$ and RI = 1.198, which corresponds to the relative (real) RI of polystyrene in seawater at 488 nm. The code was run for 999 angles (ψ) between 0–180°, resulting in a 0.180° angular resolution, and the normalized phase function (β) calculated using

$$\beta(\psi) = \frac{1}{\pi x^2 Q_{\text{sca}}} \left\{ \frac{|S_1|^2 + |S_2|^2}{2} \right\}, \tag{B1}$$

where x is the electric size of the particle, which takes into consideration the physical size (D) and the wavelength of incident light (λ); $x = \pi D/\lambda$; and Q_{sca} is the calculated Mie scattering efficiency. The second bracketed term in Eq. (B1) is the average intensity at a given angle ψ ; S_1 and S_2 are the parallel and perpendicularly scattered components, respectively, to the incident beam. A numerical integration consistency check on values calculated using Eq. (B1) was carried out to ensure that the criterion of Eq. (A8) was satisfied. A Monte Carlo simulation was carried out in order to estimate the percentage of scattered light intercepted by the forward and side detectors for each 0.180° increment in the scattering angle. A pair of random numbers was generated (R_ψ , R_ϕ) and used to compute values for ψ and ϕ , and they thus defined a scatter direction: ψ was determined

using the calculated values of $\beta(\psi)$ [Eq. (B1)] and using its cumulative distribution function characteristics [Eq. (A8)]; ϕ was determined using Eq. (A5). This procedure was repeated 10^8 times to statistically represent different photon trajectories. To determine if the individual photons impacted the forward or side detector, the angles (ψ , ϕ) were used, together with the internal measurements of the flow cytometer, to convert into Cartesian coordinates such that

$$\begin{aligned} fy &= x_{\text{lens}} \tan(\psi) \cos(\phi - \pi) \\ fz &= x_{\text{lens}} \tan(\psi) \sin(\phi - \pi) \\ fr &= x_{\text{lens}} \tan(\theta_{f2}), \end{aligned} \quad (\text{B2})$$

where fy , fz describe a forward position in the y , z plane; x_{lens} is the distance from the particle stream and the forward detector lens; fr is the forward lens radius; and θ_{f2} is the maximum acceptance angle of the forward lens. For a lens of known dimensions, the fraction of statistically generated photon trajectories falling within fr , taking into consideration the width of the forward obscuration bar (1 mm), was calculated in order to obtain the forward scattering weighting function $W_{S3}(\psi)$. Similarly, for the side scatter lens

$$\begin{aligned} sx &= y_{\text{lens}} \tan\left(\frac{\pi}{2} - \psi\right), \\ sz &= y_{\text{lens}}(\phi - \pi), \\ sr &= y_{\text{lens}} \tan\left(\frac{\pi}{2} - \theta_{s1}\right), \end{aligned} \quad (\text{B3})$$

where sx , sz describe a sideward position in the x , z plane; y_{lens} is the distance from the particle stream and the side detector lens; sr is the side detector lens radius; and θ_{s1} is the minimum angle of the side detector lens. For a lens of known dimensions, the fraction of statistically generated photon trajectories falling within sr was calculated in order to obtain the side scattering weighting function $W_{S4}(\psi)$. The final form of $W(\psi)$, specific to the internal optical geometries of the flow cytometers used in this paper, is shown in Fig. 12, which combines the two individual calculated weighting functions $W_{S3}(\psi)$ and $W_{S4}(\psi)$.

APPENDIX C: CONSTRUCTION OF THE MIE LUT

Following Bohren and Huffman [15], and in agreement with van de Hulst [45], the phase function is related to the differential scattering cross section using the relationship

$$\beta = \frac{1}{\pi x^2 Q_{\text{sca}}} \frac{dC_{\text{sca}}}{d\Omega}. \quad (\text{C1})$$

Mie scattering calculations were carried out at 0.108° resolution for particles of diameter (D) ranging between 0.02 – $10 \mu\text{m}$ ($\Delta D = 0.02 \mu\text{m}$) and real RI between 1.010 and 1.200 ($\Delta \text{RI} = 0.001$) in order to form a high-resolution LUT containing pairs of $[(dC_{\text{sca}}/d\Omega_{S3}), (dC_{\text{sca}}/d\Omega_{S4})]$. These were numerically calculated using

$$\frac{dC_{\text{sca}}}{d\Omega_{S3}} = \int_{\psi=1.5^\circ}^{\psi=12.2^\circ} \beta(\psi) W(\psi) Q_{\text{sca}} \pi r^2 d\psi, \quad (\text{C2})$$

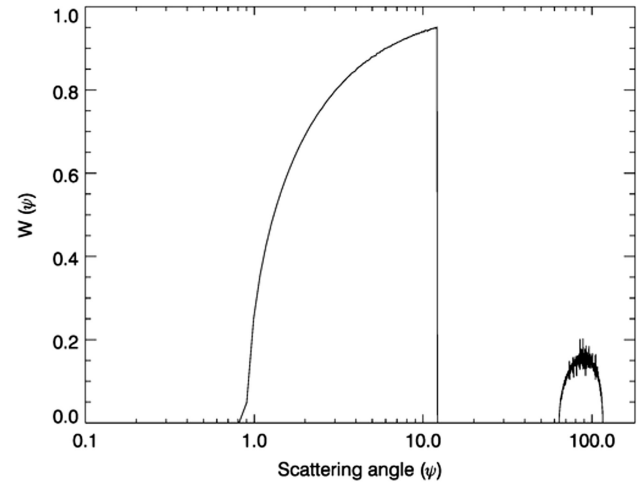


Fig. 12. Weighting function for FACSsort flow cytometer accounting for forward obscuration bar ($\psi < 1.0$) and for both forward and side scattering detector geometries (see text for details). Weighting function constructed using Mie scattering theory and a Monte Carlo approach (10^8 photons). Note log transformation of the x axis.

$$\frac{dC_{\text{sca}}}{d\Omega_{S4}} = \int_{\psi=116^\circ}^{\psi=64^\circ} \beta(\psi) W(\psi) Q_{\text{sca}} \pi r^2 d\psi, \quad (\text{C3})$$

where r is the particle radius and $d\psi$ the angular resolution of the Mie calculations. This enabled the size and RI of a particle p within the flow cytometer to be retrieved for any given pair of $[(dC_{\text{sca}}/d\Omega_{S3})_p, (dC_{\text{sca}}/d\Omega_{S4})_p]$, once the LUT had been effectively calibrated against a reference particle (or stream of identical reference particles) of known dimensions and RI. The theoretical results are presented in Fig. 2.

Funding. Natural Environment Research Council (NE/R015953/1); Simons Foundation (CBIOMES (Grant ID: 549947,SS)).

Acknowledgment. This work was funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science, and is a contribution to the Atlantic Meridional Transect and Theme 1.3 – Biological Dynamics. This work was supported by the Simons Collaboration on Computational Biogeochemical Modeling of Marine Ecosystems/CBIOMES. T.S. and G.T. thank Richard Geider, who started the whole thought process in 2001, which culminated in this paper almost two decades later. Figures 8, 10, and 11 were produced using Ocean Data View. This study contributes to the international IMBeR project and is contribution number 340 of the AMT programme..

REFERENCES

1. T. J. Smyth, G. F. Moore, T. Hirata, and J. Aiken, "Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment," *Appl. Opt.* **45**, 8116–8131 (2006).

2. Z. P. Lee, K. L. Carder, and R. A. Arnone, "Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters," *Appl. Opt.* **41**, 5755–5772 (2002).
3. C. D. Mobley, *Light and Water; Radiative Transfer in Natural Waters* (Academic, 1994).
4. C. E. Widdicombe, D. Eloire, D. Harbour, R. P. Harris, and P. J. Somerfield, "Long-term phytoplankton community dynamics in the Western English Channel," *J. Plankton Res.* **32**, 643–655 (2010).
5. S. Menden-Deuer and E. J. Lessard, "Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton," *Limnol. Oceanogr.* **45**, 569–579 (2000).
6. E. Alvarez, A. Lopez-Urrutia, E. Nogueira, and S. Fraga, "How to effectively sample the plankton size spectrum? A case study using FlowCAM," *J. Plankton Res.* **33**, 1119–1133 (2011).
7. H. H. Jakobsen and J. Carstensen, "FlowCAM: sizing cells and understanding the impact of size distributions on biovolume of planktonic community structure," *Aquat. Microb. Ecol.* **65**, 75–87 (2011).
8. J. M. Sieburth, *Sea Microbes* (Oxford University, 1979).
9. M. V. Zubkov, M. A. Sleigh, G. A. Tarran, P. H. Burkill, and R. J. G. Leakey, "Picoplanktonic community structure on an Atlantic transect from 50 degrees N to 50 degrees S," *Deep-Sea Res. Part A* **45**, 1339–1355 (1998).
10. G. A. Tarran, J. L. Heywood, and M. V. Zubkov, "Latitudinal changes in the standing stocks of eukaryotic nano- and picophytoplankton in the Atlantic Ocean," *Deep-Sea Res. Part II* **53**, 1516–1529 (2006).
11. R. W. Spinrad and J. F. Brown, "Relative real refractive index of marine microorganisms: a technique for flow cytometric estimation," *Appl. Opt.* **25**, 1930–1934 (1986).
12. S. G. Ackleson and R. W. Spinrad, "Size and refractive-index of individual marine particulates—a flow cytometric approach," *Appl. Opt.* **27**, 1270–1277 (1988).
13. R. E. Green, H. M. Sosik, R. J. Olson, and M. D. DuRand, "Flow cytometric determination of size and complex refractive index for marine particles: comparison with independent and bulk estimates," *Appl. Opt.* **42**, 526–541 (2003).
14. J. Agagliate, R. Rottgers, M. S. Twardowski, and D. McKee, "Evaluation of a flow cytometry method to determine size and real refractive index distributions in natural marine particle populations," *Appl. Opt.* **57**, 1705–1716 (2018).
15. C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles* (Wiley, 1983).
16. S. W. Chisholm, R. J. Olson, E. R. Zettler, R. Goericke, J. B. Waterbury, and N. A. Welschmeyer, "A novel free-living prochlorophyte abundant in the oceanic euphotic zone," *Nature* **334**, 340–343 (1988).
17. P. W. Johnson and J. M. Sieburth, "Chroococcoid cyanobacteria in the sea—ubiquitous and diverse phototropic biomass," *Limnol. Oceanogr.* **24**, 928–935 (1979).
18. J. B. Waterbury, S. W. Watson, R. R. L. Guillard, and L. E. Brand, "Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium," *Nature* **277**, 293–294 (1979).
19. T. Smyth, G. Quartly, T. Jackson, G. Tarran, M. Woodward, C. Harris, C. Gallienne, R. Thomas, R. A. Rees, D. Cummings, R. Brewin, V. Kitidis, J. Stephens, M. Zubkov, and A. Rees, "Determining Atlantic Ocean province contrasts and variations," *Prog. Oceanogr.* **158**, 19–40 (2017).
20. A. P. Rees, P. D. Nightingale, A. J. Poulton, T. J. Smyth, G. A. Tarran, and G. H. Tilstone, "The Atlantic Meridional transect programme (1995–2016)," *Prog. Oceanogr.* **158**, 3–18 (2017).
21. G. Walther, N. Zimmerman, W. Moore, D. Parks, S. Meehan, I. Belitskaya, J. Pan, and L. Herzenberg, "Automatic clustering of flow cytometry data with density-based merging," *Adv. Bioinform.* **2009**, 686759 (2009).
22. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (AAAI, 1996), 226–231.
23. I. Z. Kozma, P. Krok, and E. Riedle, "Direct measurement of the group-velocity mismatch and derivation of the refractive-index dispersion for a variety of solvents in the ultraviolet," *J. Opt. Soc. Am. B* **22**, 1479–1485 (2005).
24. K. Kerl and H. Varchmin, "Refractive index dispersion (RID) of some liquids in the UV/VIS between 20°C and 60°C," *J. Mol. Struct.* **349**, 257–260 (1995).
25. W. M. Haynes, *CRC Handbook of Chemistry and Physics*, 94th ed. (CRC Press, 2014).
26. R. W. Austin and G. Halikas, *The Index of Refraction of Seawater* (Scripps Institution of Oceanography, 1976).
27. G. Rocap, F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm, "Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation," *Nature* **424**, 1042–1047 (2003).
28. E. R. Zinser, A. Coe, Z. I. Johnson, A. C. Martiny, N. J. Fuller, D. J. Scanlan, and S. W. Chisholm, "Prochlorococcus ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method," *Appl. Environ. Microbiol.* **72**, 723–732 (2006).
29. J. W. Chandler, Y. J. Lin, P. J. Gainer, A. F. Post, Z. I. Johnson, and E. R. Zinser, "Variable but persistent coexistence of *Prochlorococcus* ecotypes along temperature gradients in the ocean's surface mixed layer," *Environ. Microbiol. Rep.* **8**, 272–284 (2016).
30. E. Aas, "Refractive index of phytoplankton derived from its metabolite composition," *J. Plankton Res.* **18**, 2223–2249 (1996).
31. C. M. Moore, M. M. Mills, K. R. Arrigo, I. Berman-Frank, L. Bopp, P. W. Boyd, E. D. Galbraith, R. J. Geider, C. Guieu, S. L. Jaccard, T. D. Jickells, J. La Roche, T. M. Lenton, N. M. Mahowald, E. Maranon, I. Marinov, J. K. Moore, T. Nakatsuka, A. Oschlies, M. A. Saito, T. F. Thingstad, A. Tsuda, and O. Ulloa, "Processes and patterns of oceanic nutrient limitation," *Nat. Geosci.* **6**, 701–710 (2013).
32. A. Longhurst, *Ecological Geography of the Sea* (Academic, 1998).
33. H. W. Balfourt, T. Berman, S. Y. Maestrini, A. Wenzel, and T. Zohary, "Flow-cytometry—instrumentation and application in phytoplankton research," *Hydrobiologia* **238**, 89–97 (1992).
34. F. Partensky, W. R. Hess, and D. Vaulot, "Prochlorococcus, a marine photosynthetic prokaryote of global significance," *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).
35. J. B. Waterbury, J. M. Willey, D. G. Franks, F. W. Valois, and S. W. Watson, "A cyanobacterium capable of swimming motility," *Science* **230**, 74 (1985).
36. Q. Fu, T. J. Thorsen, J. Su, J. M. Ge, and J. P. Huang, "Test of Mie-based single-scattering properties of non-spherical dust aerosols in radiative flux calculations," *J. Quant. Spectrosc. Radiat. Transfer* **110**, 1640–1653 (2009).
37. E. Organelli, G. Dall'Olmo, R. J. W. Brewin, G. A. Tarran, E. Boss, and A. Bricaud, "The open-ocean missing backscattering is in the structural complexity of particles," *Nat. Commun.* **9**, 5439 (2018).
38. A. Quirantes and S. Bernard, "Light-scattering methods for modelling algal particles as a collection of coated and/or nonspherical scatterers," *J. Quant. Spectrosc. Radiat. Transfer* **100**, 315–324 (2006).
39. D. Stramski, "Refractive index of planktonic cells as a measure of cellular carbon and chlorophyll a content," *Deep-Sea Res. Part I* **46**, 335–351 (1999).
40. M. D. DuRand, R. E. Green, H. M. Sosik, and R. J. Olson, "Diel variations in optical properties of *micromonas pusilla* (prasinophyceae)," *J. Phycology* **38**, 1132–1142 (2002).
41. J. Agagliate, R. Rottgers, K. Heymann, and D. McKee, "Estimation of suspended matter, organic carbon, and chlorophyll-a concentrations from particle size and refractive index distributions," *Appl. Sci.* **8**, 2676 (2018).
42. E. A. Laws, "Photosynthetic quotients, new production and net community production in the open ocean," *Deep-Sea Res. Part I* **38**, 143–167 (1991).
43. D. Woitzik, J. Weckesser, and U. J. Jurgens, "Isolation and characterization of cell-wall components of the unicellular cyanobacterium *synechococcus* sp pcc 6307," *J. Gen. Microbiol.* **134**, 619–627 (1988).
44. A. C. Redfield, "The biological control of chemical factors in the environment," *Am. Sci.* **46**, 205–221 (1958).
45. H. C. van de Hulst, *Light Scattering by Small Particles* (Dover, 1957).