Supplement to "Ocean Color Algorithm for the Retrieval of the Particle Size Distribution and Carbon-Based Phytoplankton Size Classes Using a Two-Component Coated-Spheres Backscattering Model"

Tihomir S. Kostadinov et al.

S1 Assessment and propagation of uncertainties

The Monte Carlo simulations (N=3000) of the two-population Mie scattering model were used to estimate uncertainty of the retrieved PSD parameters that is due to assumptions of the values of the model input parameters. The input distributions from which the varying input parameters were sampled for the Monte Carlo simulations are shown in main text Tables 1 and 2. This approach resulted in 3000 different realizations of modeled spectral b_{bp}, for each ξ input value. The median of all realizations is used as the SAM end-members, and Kruskal-Wallis analysis of variance is used to determine class similarity (where each input ξ value represents a class). This analysis determined, for each class, the neighboring classes that are statistically not different from it (at the 5% significance level) in terms of spectral angle. This gave a range of possible retrieved ξ values. This range is not always symmetric, but it is approximated as a symmetric 95% confidence interval (CI) and converted to standard deviation for use in subsequent analytical error propagation. The standard deviation of the corresponding N_0 parameter across all similar classes and all Monte Carlo realizations is used as the uncertainty estimate for N_0.

The uncertainties of the PSD parameters and allometric coefficients were used with the first order derivatives of the derived products (absolute and fractional size-partitioned phyto C) with respect to those variables to analytically propagate the uncertainty estimate to the derived products. Phyto C absolute values are functions of both PSD parameters and both allometric coefficients, whereas the fractional values are functions only of ξ and the b allometric coefficient because a ratio of phyto C values is taken (Eq. 5). We use the allometric coefficients obtained by Roy et al. (2017), who use information from three different regressions presented in Menden-Deuer and Lessard (2000). The weighted mean of the confidence intervals of the Menden-Deuer and Lessard (2000) allometric coefficients was used. Uncertainties were also propagated analytically to the composite imagery (monthly and overall climatologies) provided in the OC-CCI-derived PSD/phyto C data set (linked in main text Sec. 4) For more details see the scientific code used to do the uncertainty propagation (linked in main text Sec. 4), as well as Kostadinov et al. (2016) and their Supplement.

S2 Modeled Backscattering Efficiencies

Example output phytoplankton Q_{bb} are plotted in Fig. S2A. For a fixed diameter, wavelength can significantly affect the values, especially at intermediate diameters. Coated sphere scattering patterns also exhibit spikes, the influence of which on computed IOPs is minimized here by having a high sampling frequency in diameter, D. Figure S2A also shows the Q_{bb} curves corresponding to the Gladstone-Dale (GD) equivalent homogeneous spheres. For most diameters, homogeneous spheres exhibit significantly lower backscattering efficiencies, which is consistent with prior studies (e.g. Bernard et al. (2009); Quirantes and Bernard (2006); Organelli et al. (2018)). This indicates that the contribution of phytoplankton to backscattering in the global ocean has likely been underestimated in previous studies (see Dall'Olmo et al. (2009)), in which homogeneous spheres were assumed for modeling of phytoplankton (e.g. Stramski and Kiefer (1991); Kostadinov et al. (2009)).

The difference between modeled phytoplankton and NAP Q_{bb} curves is shown in Fig. S2B. Unlike panel A, the curves here are band-averaged (see main text Sec. 2.2) and the median across all Monte Carlo runs is shown as an example. Note that NAP are modeled over a wider range of diameters as compared with phytoplankton. Phytoplankton represented as coated spheres still exhibit higher efficiencies than homogeneous NAP, but mostly only over the 1 to 10 $\mu\mathrm{m}$ diameter range. The difference here is driven by the choices of Mie inputs for phytoplankton vs. NAP. In particular, NAP are allowed to have a wider range of real RIs, accommodating possible contributions by minerogenic particles. Purely organic detrital NAP would have lower Q_{bb} than is shown in Supplementary Fig. S2B.

S3 Cumulative Backscattering Analysis

Here, we focus on investigating what particle size ranges contribute to the backscattering signal, which is investigated via cumulative b_{bp} plots (Fig. S4). Cumulative modeled b_{bp} curves for phytoplankton (Fig. S4A, 443 and 560 nm shown) indicate that, as expected, the higher the PSD slope, the larger fraction of the phytoplankton b_{bp} signal is due to the smallest phytoplankton cells. Generally, the 443 and 560 nm bands behave similarly, however, differences exist, e.g. at PSD slopes 3.5 and 4.0; specifically, at ξ=4.0, $\approx$60% of the signal is due to phytoplankton cells with D < 3 $\mu\mathrm{m}$ at 560 nm, whereas this value rises to $\approx$85% at 443 nm. A sigmoidal-like shape of this curve indicates that asymptotic values tend to be reached at low diameters and high diameters, i.e. most of the b_{bp} signal is captured within the range of D values used. Note that the size of the smallest photoautotroph - *Prochlorococcus* ($\approx$ 0.5 $\mu\mathrm{m}$ in diameter (Morel et al. (1993)) informs the choice of the lower diameter limit used for phytoplankton (Table 1).

Cumulative NAP b_{bp} for modeled spectra (Supplement Fig. S4B) indicates that for most PSD slopes, the size range used captures most of the signal, if NAP can be reasonably represented by homogeneous spheres. Importantly, for ξ=4.0, most of the signal comes from particles with D between 0.1 and 1.0 $\mu\mathrm{m}$, and these smaller particles are also essential for reproducing modeled spectral shapes in the same range as observed satellite b_{bp} spectra. Coated spheres representing phytoplankton within their size range alone are not able to do so (not shown). This is an important observation, illustrating the need to better understand the nature of sub-micron particles (e.g. Stramski et al. (2004)). The difference between coated spheres and their Gladstone-Dale (GD) homogeneous equivalents in terms of cumulative b_{bp} (Fig. S4C) are a function of the differences in Q_{bb}

curve shapes (Fig. S2A) and the PSD slope used. At intermediate values of the PSD slope such as 4.0, ≈30% of the signal is due to particles smaller than D = 1.0 μm in the case of coated spheres, whereas almost 60% of the signal is due to particles smaller than this threshold in the case of homogeneous spheres.

S4 Details of Empirical Tuning Derivation

This section describes the details of the derivation of the N_O empirical tuning method. Monthly OC-CCI v5.0 imagery for years 2004 and 2015 (48 total images, 24 each for Chl and POC) were used to compute the PSD-derived Chl and POC. These values were then compared on a per-pixel basis against the standard OC-CCI v5.0 Chl product, and against the Stramski et al. (2008) POC retrieval (termed the standard values here). These comparisons were used in an optimization procedure that finds the best linear regression correction of the decimal logarithm of N_0. The cost function used was the average absolute value of the decimal logarithm of the ratio between the PSD-based and the standard values, summing the Chl and POC contributions and weighting them equally. This cost function is minimal when both Chl and POC standard and PSD-based retrievals are the same everywhere. The optimization was performed using the $\mathrm{Statistics\ and\ MachineLearning}$® Toolbox of MATLAB®. A hybrid optimization technique was implemented, using genetic algorithms (ga function) and the $\mathrm{fmincon}$ function. The optimization for each month returns one global value of linear slope and intercept to be applied to the original, un-tuned N_0 retrievals, in order to obtain the tuned retrievals (henceforth retrievals without tuning applied are referred to as "original"). The median of the 48 tuning slopes and intercepts thus obtained were subsequently used to adjust N_0 in log10 space as follows:

S5 Supplement Figures

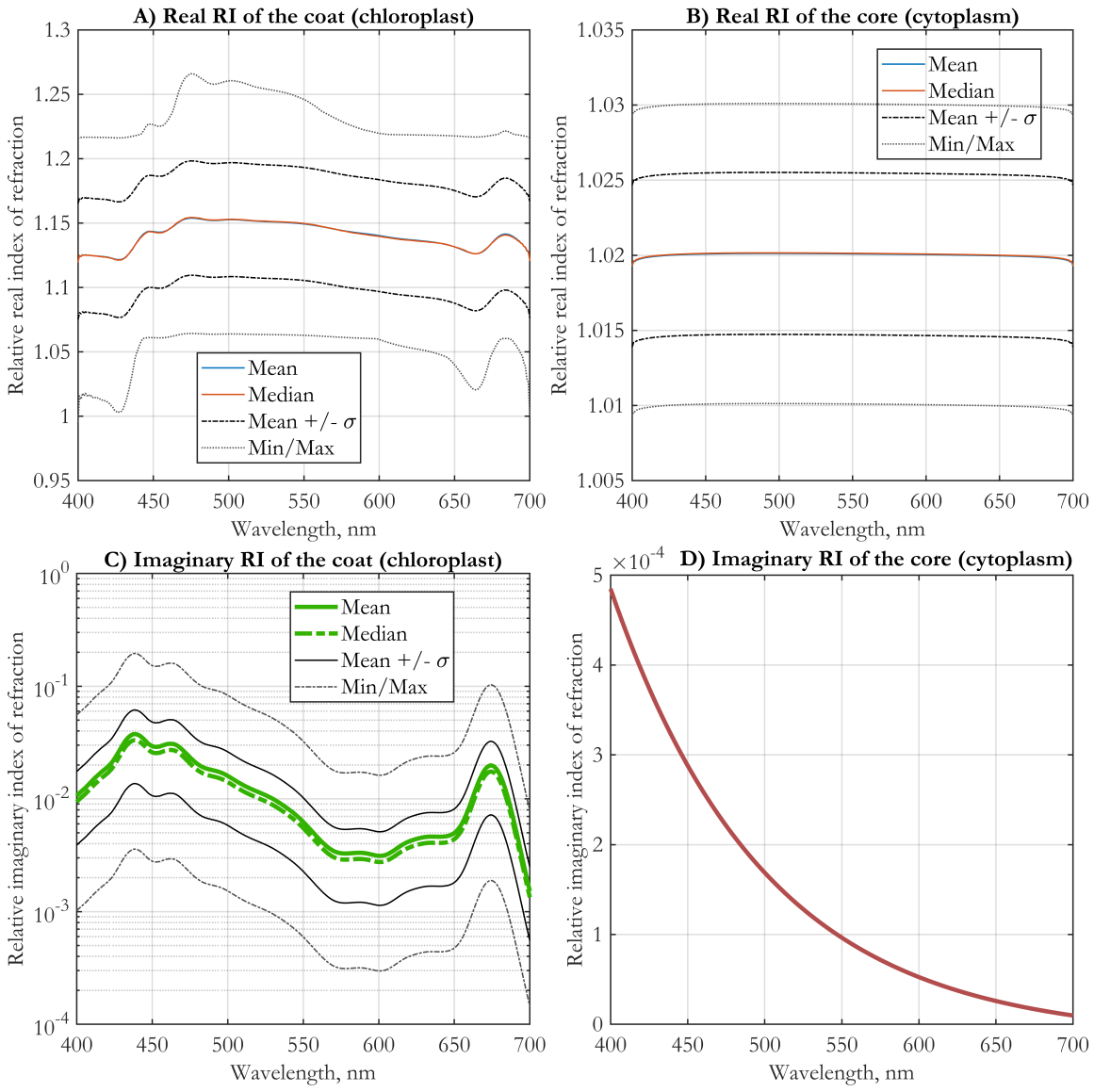

Figure S1. EAP-based hyperspectral complex refractive index (RI) inputs to the coated spheres (representing phytoplankton cells) Mie code, given relative to the index of refraction of seawater: A) real RI of the coat as chloroplast, B) real RI of the core as cytoplasm, C) imaginary RI of the coat as chloroplast, and D) imaginary RI of the core as cytoplasm. In panel A, the mean and median curves almost coincide. For panels A-C, the statistics of the input distributions used in the Monte Carlo simulation are given, namely mean, median, mean plus and minus one standard deviation, and the minimum and maximum values. Panels A and B exhibit spectral dependencies according to the Kramers-Kronig relations.

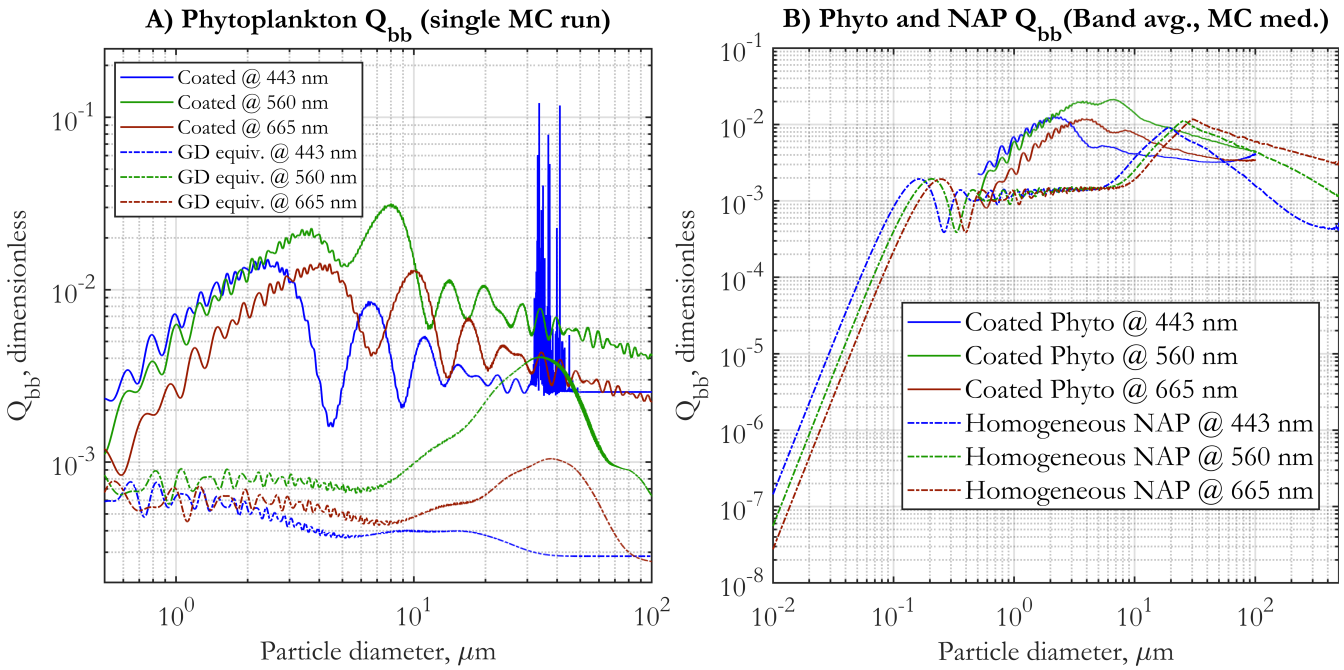

Figure S2. (A) Example backscattering efficiencies, Q_{bb}, for phytoplankton modeled as coated spheres. A single Aden-Kerker computation run is shown, using the median inputs from the Monte Carlo simulations. The coated sphere results for three different wavelengths (corresponding to OLCI nominal bands and the OC-CCI v5.0 data set bands) are shown as solids lines. The corresponding Gladstone-Dale (GD) - equivalent homogeneous sphere results are shown in dash-dot lines. (B) Band-averaged backscattering efficiencies, $\overline{Q_{bb}}$, for phytoplankton modeled as coated spheres (solid lines) and NAP modeled as homogeneous spheres (dash-dot lines). The median values across all Monte Carlo simulation runs are shown.

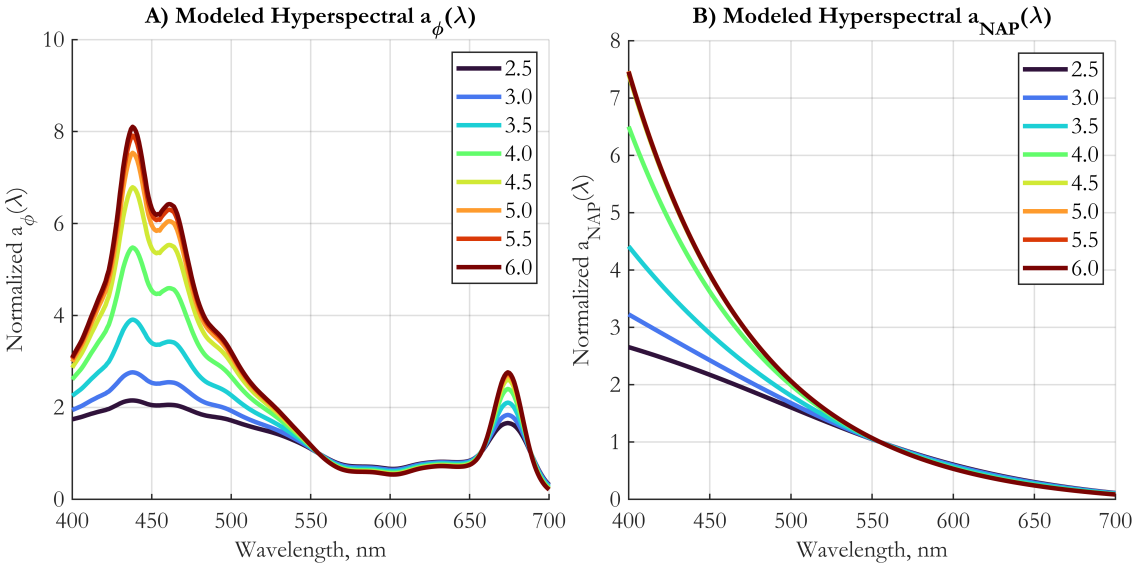

Figure S3. Modeled hyperspectral particulate absorption coefficient by (A) phytoplankton, $a_\phi(\lambda)$, using EAP-based coated spheres Mie scattering computations, and (B) NAP,$a_{NAP}(\lambda)$, modeled as homogeneous spheres, as a function of the input power-law PSD slope (color-coded solid lines, as in legend). All spectra are shown normalized to the respective values at 555 nm.

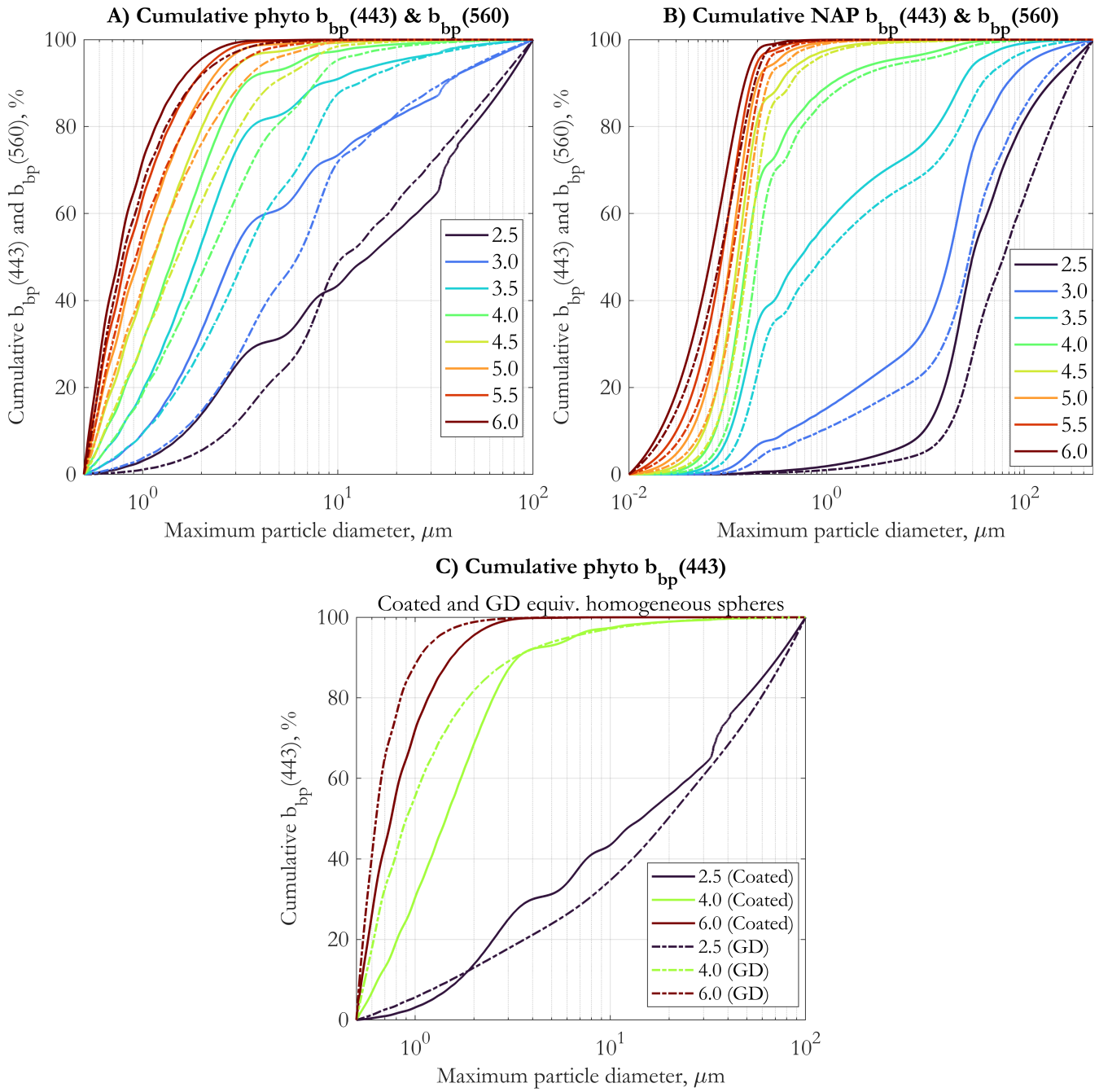

Figure S4. Cumulative $b_{bp}(443)$ (solid lines) and $b_{bp}(560)$ (dash-dot lines) for phytoplankton (A) and NAP (B) as a function of PSD slope ξ (color coded as in legend). For each particle diameter (x-axes) the plots show the percentage of backscattering due to particles smaller than or equal to that diameter (y-axes). (C) Cumulative $b_{bp}(443)$ for phytoplankton modeled as coated spheres (solid lines) and for the corresponding Gladstone-Dale (GD) equivalent homogeneous spheres (dash-dot lines). Curve pairs are shown for three different PSD slopes ξ, namely 2.5, 4.0 and 6.0 (color coded as in legend). All panels display results of the single forward illustrative Mie calculation run which uses medians of the Monte Carlo varied inputs.

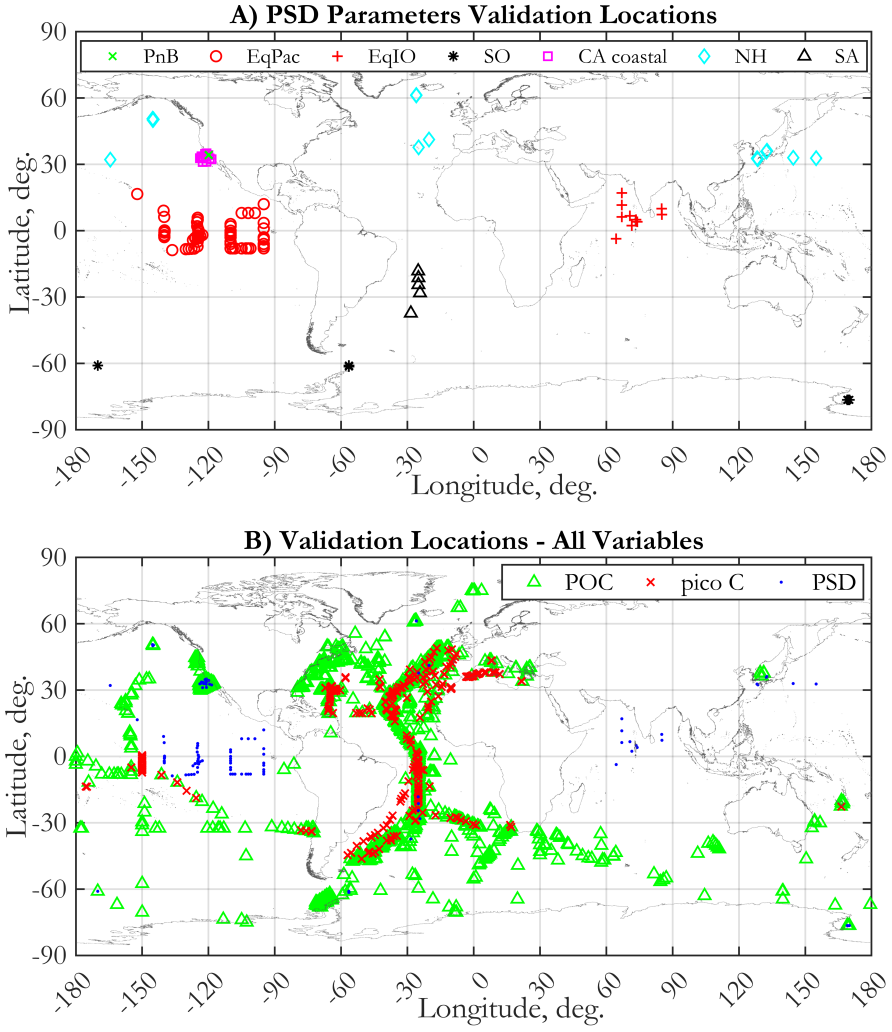

Figure S5. A) Locations of *in-situ* PSD data used in algorithm validation. Different regions are shown using symbols as in the legend, in which acronyms are as follows - Plumes and Blooms Project (PnB), Equatorial Pacific (EqPac), Equatorial Indian Ocean (EqIO), Southern Ocean (SO), CA coastal areas excluding PnB (CA coastal), Northern Hemisphere higher latitudes (NH), and Southern Atlantic (SA). B) Locations of *in-situ* data used in validation match-ups for all variables, as shown in the legend - particulate organic carbon (POC, green triangle), pico-phytoplankton carbon (pico C, red 'x'), and PSD (blue dot).

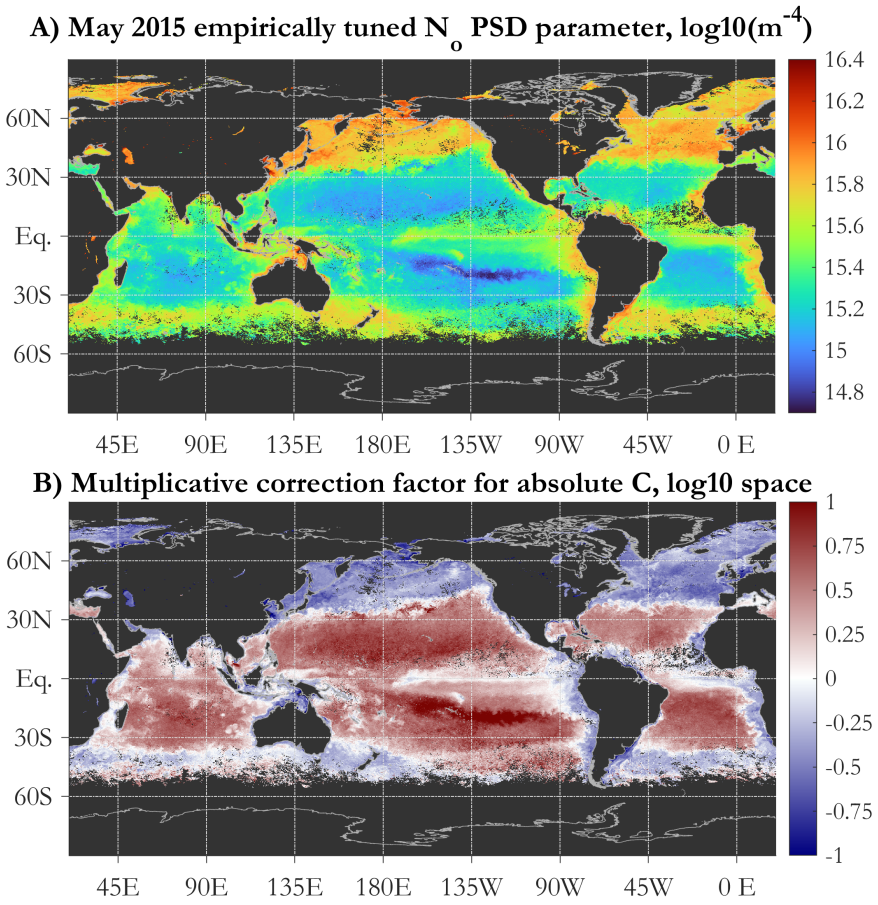

Figure S6. (A) The N_0 parameter (m^{-4}, mapped in log10 space) after tuning has been applied (main text Eq. 7) in order to achieve more realistic global estimates of POC and Chl via the PSD model presented here; (B) A multiplicative tuning factor (dimensionless) to be applied to absolute carbon (and Chl) estimates from the PSD-based algorithm; the factor is applied in linear space, but it is plotted in log10 space, i.e. a map value of 0 indicates a tuning factor of 1, meaning that the tuned and original values are the same. Both panels use the monthly OC-CCI v5.0 data for May 2015 as input.

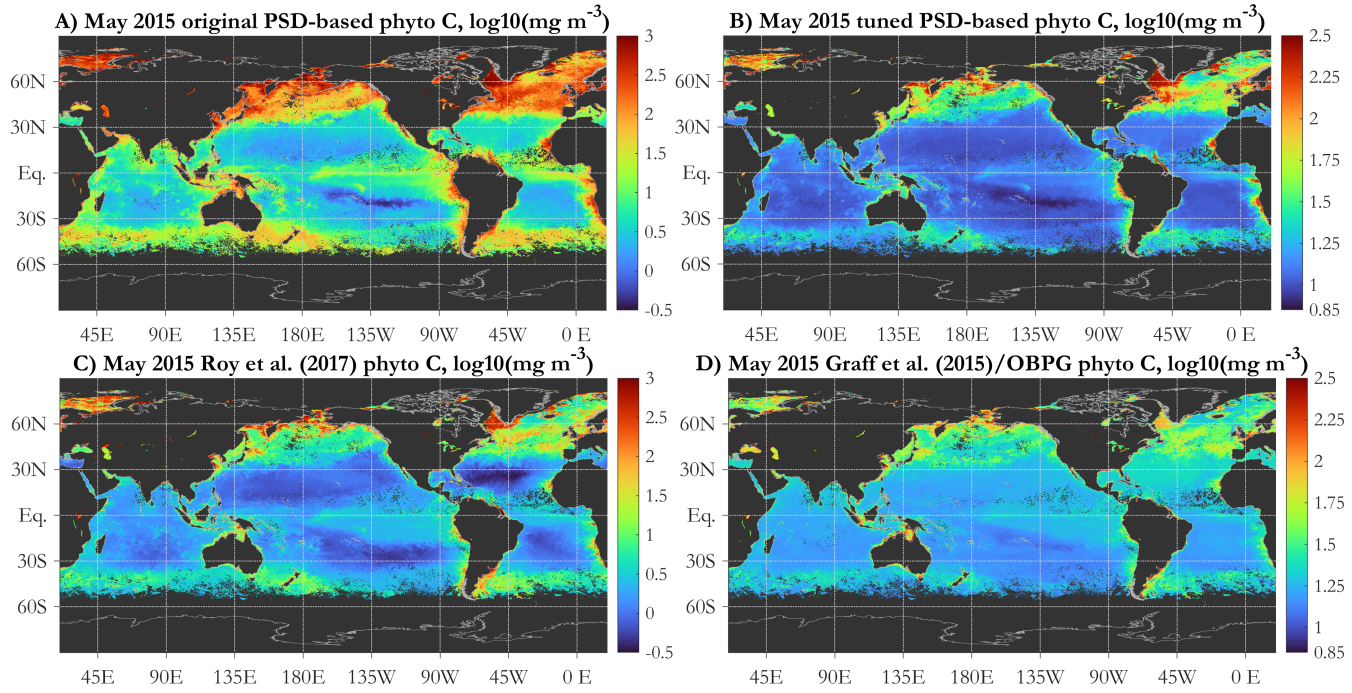

Figure S7. (A) Original PSD-based phyto C, same as main text Fig. 4A; (B) Tuned PSD-based phyto C; (B) The Roy et al. (2017) absorption-based and PSD-based phyto C retrieval, and (C) the Graff et al. (2015) phyto C retrieval, based on a scaling of b_{bp}. All three panels use the monthly OC-CCI v5.0 data for May 2015 as input. Note the different color scale of panels A and C vs. panels B and D.

References

Bernard, S., Probyn, T. A., and Quirantes, A.: Simulating the optical properties of phytoplankton cells using a two-layered spherical geometry, Biogeosciences Discussions, 6, 1497–1563, https://doi.org/10.5194/bgd-6-1497-2009, 2009.

Dall'Olmo, G., Westberry, T. K., Behrenfeld, M. J., Boss, E., and Slade, W. H.: Significant contribution of large particles to optical backscattering in the open ocean, Biogeosciences, 6, 947–967, https://doi.org/10.5194/bg-6-947-2009, 2009.

Graff, J. R., Westberry, T. K., Milligan, A. J., Brown, M. B., Dall'Olmo, G., van Dongen-Vogels, V., Reifel, K. M., and Behrenfeld, M. J.: Analytical phytoplankton carbon measurements spanning diverse ecosystems, Deep-Sea Research Part I: Oceanographic Research Papers, 102, 16–25, https://doi.org/10.1016/j.dsr.2015.04.006, 2015.

Kostadinov, T., Siegel, D., and Maritorena, S.: Retrieval of the particle size distribution from satellite ocean color observations, Journal of Geophysical Research: Oceans, 114, https://doi.org/10.1029/2009JC005303, 2009.

Kostadinov, T., Milutinovic, S., Marinov, I., and Cabré, A.: Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution, Ocean Science, 12, https://doi.org/10.5194/os-12-561-2016, 2016.

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

Morel, A., Yu-Hwan Ahn, Partensky, F., Vaulot, D., and Claustre, H.: Prochlorococcus and Synechococcus: a comparative study of their optical properties in relation to their size and pigmentation, Journal of Marine Research, 51, 617–649, https://doi.org/10.1357/0022240933223963, 1993.

Organelli, E., Dall'Olmo, G., Brewin, R. J., Tarran, G. A., Boss, E., and Bricaud, A.: The open-ocean missing backscattering is in the structural complexity of particles, Nature Communications, 9, https://doi.org/10.1038/s41467-018-07814-6, 2018.

Quirantes, A. and Bernard, S.: Light-scattering methods for modelling algal particles as a collection of coated and/or nonspherical scatterers, Journal of Quantitative Spectroscopy and Radiative Transfer, 100, 315–324, https://doi.org/10.1016/j.jqsrt.2005.11.048, 2006.

Roy, S., Sathyendranath, S., and Platt, T.: Size-partitioned phytoplankton carbon and carbon-to-chlorophyll ratio from ocean colour by an absorption-based bio-optical algorithm, Remote Sensing of Environment, 194, 177–189, https://doi.org/10.1016/j.rse.2017.02.015, 2017.

Stramski, D. and Kiefer, D. A.: Light scattering by microorganisms in the open ocean, Progress in Oceanography, 28, 343–383, https://doi.org/10.1016/0079-6611(91)90032-H, 1991.

Stramski, D., Boss, E., Bogucki, D., and Voss, K. J.: The role of seawater constituents in light backscattering in the ocean, Progress in Oceanography, 61, 27–56, https://doi.org/10.1016/j.pocean.2004.07.001, 2004.

Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R., Sciandra, A., Stramska, M., Twardowski, M. S., Franz, B. A., and Claustre, H.: Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans, Biogeosciences, 5, 171–201, https://doi.org/10.5194/bg-5-171-2008, 2008.