Supplement of Ocean Sci., 12, 561–575, 2016
http://www.ocean-sci.net/12/561/2016/
doi:10.5194/os-12-561-2016-supplement

Ocean Science Open Access EGU

Supplement of

Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution

Tihomir S. Kostadinov et al.

Correspondence to: Tihomir S. Kostadinov (kostadinov.t@gmail.com)

Supplement

to Kostadinov, Milutinović, Marinov and Cabré, "Carbon-Based Phytoplankton Size Classes Retrieved via Ocean Color Estimates of the Particle Size Distribution"

S1. Additional Methodology Information

S1.1. Details of Allometric Relationships Application

There is some variation in the literature regarding the integration limits of Eq. 4 and the cut-off values that are used in definition of the PSCs, because they are rather arbitrary. The only objective cut-off value is that for the minimum autotrophic picoplankton size (0.5 μm), as this is the reported ESD of the smallest known marine photosynthesizer (*Partensky et al., 1999*). The maximum size threshold for unicellular phytoplankton is not as clear and we settle on 50 μm, since larger algal cells are seldom encountered even in eutrophic conditions (*Charles Stock, personal communication, 2013*) and are particularly rare in the open ocean (*Roy et al.*, *2013*).

The parameters *a* and *b* in Eq. 4, pertaining to the groups of phytoplankton that are relevant for this study, are selected from *MDL2000* and are presented in Table S1. C biomass is estimated using more than one allometric relationship in order to achieve a globally applicable solution. The biomass of picoplankton is computed by implementing the parameters for cells with volume below 3000 μm^3 (D < 17.894 μm) in Eq. 4 for *D* from 0.5 μm to 2 μm. Nanoplankton biomass is computed by combining all the three sets of *a* and *b* values listed in Table S1. The first set (same as for picoplankton) is used in Eq. 4 for *D* between 2 μm and 17.894 μm. The second (for generic non-diatomaceous phytoplankton) and third set (for diatoms above 3000 μm^3) are applied separately for *D* between 17.894 μm and 20 μm and the results are averaged and added to the result of applying the first set of parameters, obtaining the overall nanoplankton biomass. Similarly, microplankton biomass is determined as the arithmetic mean of the respective output of Eq. 4 when the second and the third sets of *a* and *b* values are used for *D* from 20 μm to 50 μm. The biomass of the entire phytoplankton community (0.5 μm ≤ *D* ≤ 50 μm) is the sum of the respective biomass values for the three PSCs.

Analytical solution of the integral of Eq. 4 thus results in the following expression for a given size class:

$$C = \sum_{i=1}^{p} w_i \frac{1}{3} 10^{-9} a_i \left(\frac{10^{18}\pi}{6} \right)^{b_i} N_o D_o^{\xi} \frac{1}{3b_i - \xi + 1} \left(D_{\max i}^{3b_i - \xi + 1} - D_{\min i}^{3b_i - \xi + 1} \right) \quad \text{(S1)}$$

In the above equation, *p* represents the number of distinct sets of allometric coefficients used, i.e. p = 3 for total carbon and nanoplankton, *p* = 1 for picoplankton, and *p* = 2 for microplankton. Table S2 lists the weights w_i applied for each allometric relationship. The D_{max} and D_{min} values are selected as appropriate from the size ranges of the size class or the limits of applicability of the *i*th allometric relationship (Table S2). Eq. S1 is not valid when the denominator is exactly 0. In the very few cases when this happens operationally to within double floating point precision, the value of the PSD slope is nudged by a very small

value (much smaller than its uncertainty). The resulting operational allometric relationship is plotted in Fig. S1.

S1.2 Phytoplankton Carbon Estimates from Earth System Models

Phytoplankton carbon was also derived from the output of a group of Earth System simulations from the recent Coupled Model Intercomparison Project CMIP5 (*Taylor et al., 2012*). CMIP5 model output was downloaded from http://pcmdi9.llnl.gov/esgf-web-fe/. The models and their basic characteristics are summarized in Table S3. The marine biogeochemical routine for models CanESM2 and MRI-ESM1 is based on the basic NPZD (Nutrient Phytoplankton Zooplankton Detritus) structure with only one phytoplankton type and one nutrient (nitrate). The complexity increases with MPI-ESM, NorESM1, HadGEM2, and GISS-E2 via inclusion of more nutrients (nitrate, silicate, iron) and additional types of phytoplankton for HadGEM2 and GISS-E2. Finally, IPSL-CM5, GFDL-ESM2, and CESM1-BGC are the most ecologically complex models, with at least 2 types of phytoplankton, zooplankton types, more than 20 biogeochemical tracers, and inclusion of ballast in the last two models.

We derive the ensemble mean phytoplankton C from 21 years of "present" historical output (1990 to 2010) of the variable 'phyc' ("total phytoplankton carbon concentration"). Molar concentration provided by the models (mol C m^{-3}) was converted to mass concentration (mg C m^{-3}) using the atomic weight of carbon (12.011 g/mol, *Wieser et al., 2013*). The "present" output is mostly based on the historical scenario (years 1850 to 2005) forced by observed atmospheric changes (both anthropogenic and natural). The last five years (2006 to 2010) of the "present" output are based on the RCP8.5 scenario. We selected 14 models with different resolutions (ocean grid varies from 0.5º to 2º) and complexities in their biogeochemical and ecological modules, as described in Table S3. All model output was resampled to a 1^{o} grid before calculating first the temporal average of each model individually, and then averaging each model's climatology to obtain the ensemble mean model climatology. Because of significant similarities between model pairs (*Cabré et al., 2014*), when computing ensemble averages we used weights as in Table S3. Before computing averages, biomass values below 0 were set to missing data, and in the case of the MRI-ESM1 model values below 0.01 mg m^{-3} C were also set to missing values. Those occur primarily along the coasts and are considered a numerical artifact (most are ~ 10^{-18} mg m^{-3} C in areas where biomass is expected to be high).

S1.3. In-situ POC-PSD Closure Analysis

In-situ closure (i.e. agreement) between POC and PSD data was investigated as verification (proof-of-concept) of the allometric methodology presented here. Nearly coincident observations of both PSD (Coulter Counter measurements) and POC (analytical chemical determinations) from Atlantic Meridional Transect (AMT) cruises 2, 3 and 4, conducted in 1996 and 1997, were obtained from the British Oceanographic Data Centre (BODC, http://www.bodc.ac.uk/). The 2 – 20 μm diameter range of the PSD data was used to fit a regression line on the log10-transformed, bin-width normalized data, yielding estimates of the PSD parameters, ξ and N_o. These were used as inputs to Eq. S1 to estimate allometric phytoplankton C from the PSD data. Chemical POC data were provided in units of μmol/L, which were converted to mg m^{-3} using carbon's atomic weight of 12.011 g/mol (*Wieser et al., 2013*). The nominal pore size of GF/F filters used in field measurements of

POC (e.g. *S08; Kostadinov et al., 2012*) is 0.7 μm, and the water was pre-filtered with a pore size of 200 μm, so only the 0.7 μm to 200 μm fraction is measured as POC. Phytoplankton C was then estimated from POC by multiplication by 1/3. Match-ups were then constructed between the two methods of estimating phytoplankton carbon concentration, considering two data points a valid match-up only if they were closer than 4.24 km from each other (diagonal of a 3x3 km box), samples were taken within 3 hours of each other, and within 15 m vertical separation. Using these criteria 44 match-ups were obtained.

S1.4. Propagation of Uncertainty to Carbon Products and Composite Imagery

The proximal input parameters of the absolute and fractional C-based PSC algorithm are the PSD slope ξ, the N_o parameter, and up to six allometric coefficients (Tables S1 and S2). Uncertainties (in terms of standard deviation) in these input parameters are propagated to the algorithm products on a per-pixel basis. The uncertainty of absolute or fractional carbon concentration, *C*, in any size class is estimated as

$$\sigma_C = \sqrt{\left(\frac{\partial C}{\partial \xi}\right)^2 \sigma_\xi^2 + \left(\frac{\partial C}{\partial N_o}\right)^2 \sigma_{N_o}^2 + \sum_{i=1}^{p}\left(\frac{\partial C}{\partial a_i}\right)^2 \sigma_{a_i}^2 + \sum_{i=1}^{p}\left(\frac{\partial C}{\partial b_i}\right)^2 \sigma_{b_i}^2} \quad \text{(S2)}$$

This is the standard analytical approximation of error propagation formulation (e.g. *Ku, 1966*). The partial derivatives of *C* with respect to the input parameters are calculated analytically from Eq. S1, where *p* = 1, 2 or 3 depending on the size class (Tables S1 and S2, Sect. S1.1). The KSM09 algorithm provides standard deviations of the output PSD parameters as a quantification of partial uncertainty. The *MDL2000* allometric coefficients are derived from linear regressions and their 95% confidence intervals are provided. These were converted to standard deviations by dividing by the respective cumulative t-distribution value for each case (Tables S1 and *MDL2000*, their Table 4). Note that covariances among the input parameters are ignored in Eq. S2, which can lead to under- or overestimation depending on the signs of the covariance and the derivatives. The estimate in Eq. S2 represents only a part of the uncertainty in *C*, because only parts of the PSD parameters' uncertainties are quantifiable and provided by KSM09. Unquantified sources of error are discussed qualitatively in Sect. 3.6 and S6.

Monthly and overall mission temporally averaged (composite) imagery was computed from the respective monthly maps of the carbon-based products as the arithmetic mean (in linear space) of all available monthly data for a given pixel. Uncertainties of the composite imagery data were determined as

$$\sigma_{composite} = \frac{\sqrt{\sum_{k=1}^{N} \sigma_k^2}}{N} \quad \text{(S3)}$$

where σ_k is the standard deviation for the *k*th month of the complete time series, out of a total of *N* available months. See Fig. S4 for a map of *N*. The process of averaging when producing composite imagery further reduces random errors (Eq. S3), but not consistent

bias (e.g. *Milutinović and Bertino, 2011*). Therefore, uncertainties in a single image produced from the input parameters are qualitatively different from uncertainties in a composite image produced by averaging multiple carbon product images with individually propagated errors. In this work we produce PSD-based products from monthly SeaWiFS imagery and propagate errors to each monthly image. The errors in any composite imagery are then calculated from the errors of the individual images participating in the averaging. Absolute uncertainties are discussed here in terms of one standard deviation, in the same units as the variable.

Averaging for composite imagery was not weighted by the inverse of the variance (σ^{-2}) in composite imagery as it was done by *Maritorena et al. (2010)* (their Eq. 2) so as to not bias the data to lower values. This is because this weighting is only appropriate when the measurements from the same underlying random value distribution are made (i.e. on a spatio-temporal scale on which the ocean is not expected to change intrinsically, S. Maritorena, pers. comm.), and further because our error structure is such that error values are proportional to the value of the retrieved parameter (the latter is especially true for the absolute carbon retrievals, and much less so for the fractional PSCs).

The C products are not linear in the PSD slope (Eq. S1) or the underlying backscattering coefficients and $R_{rs}(\lambda)$ values. The absolute C concentrations are linear in N_o, but not in the logarithm of N_o, and N_o is averaged temporally in log-space. Since various steps of the algorithm are not linear, calculating the C products from a composite image of $R_{rs}(\lambda)$ or the PSD parameters and calculating them from individual imagery and then averaging are not equivalent. This work uses monthly SeaWiFS maps, but in the future processing is planned from daily imagery in order to address this issue.

S1.5. Algorithm Output Analyses and Ancillary Data

In order to investigate relationships of the novel C-based products with Chl concentration, monthly mapped SeaWiFS 9 km OC4v6 Chl [mg m^{-3}] (*O'Reilly et al., 2000*), was obtained from NASA OBPG (the Ocean Biology Distributed Active Archive Center (OB.DAAC)) (reprocessing R2010.0). The mission composite was also obtained in order to study climatological relationships. The mission composite Chl image was downsampled to 1° resolution using 2D convolution and the 0.08 mg m^{-3} isoline of Chl was extracted in order to delineate the subtropical gyres on maps. For comparison purposes, phytoplankton C biomass was also estimated using the *B05* method with the same LAS2006-derived $b_{bp}(443)$ as used in the PSD-based algorithm. POC was retrieved using the *S08* algorithms (using the $R_{rs}(490)/R_{rs}(555)$ band ratio parameterized with all data, see their Table 2). The same R2010.0 SeaWiFS reflectances were used as for our algorithm. The POC retrievals were multiplied by 1/3 to approximate the living fraction.

The assumptions of the C biomass algorithm are more likely to be violated in shallow coastal regions, where non-biogenic particles may contribute to backscattering significantly (e.g. *Toole et al., 2001; Otero and Siegel, 2004*). We thus excluded the continental shelves from some analyses as indicated. The shelves were identified as areas shallower than 200 m and determined using the bathymetry data from the NOAA/NGDC ETOPO1 data set (*Amante and Eakins, 2009*), downsampled to 9 km or 1° resolutions as needed. In order to

display the coastline on maps, the L1 layer of the GSHHG v2.2.3 (*Wessel and Smith, 1996*) coastline data set was extracted with the NOAA/NGDC GEODAS-NG software.

In order to estimate global phytoplankton C biomass standing stock within the mixed layer, monthly mixed layer depth (MLD) for the 1997-2010 SeaWiFS mission period was computed from the UK Met Office Hadley Centre's monthly global objective analyses fields of seawater potential temperature and salinity (version EN3_v2a) (*Ingleby and Huddleston, 2007*). The fields were provided on a regular grid of 1° longitude/latitude resolution and 42 unequally spaced depth levels. For each grid cell and depth, seawater density was computed using the equation of state of seawater (*UNESCO et al*., 1980). Linear interpolation was then applied to every respective vertical profile of temperature and density to compute MLD, using the threshold approach of *de Boyer Montégut et al. (2004)* (±0.2°C or 0.03 kg m^{-3}). Following their recommendation, the shallower of the temperature- and density-based values was chosen as the best estimate of MLD. Monthly and overall SeaWiFS-era global MLD composites were computed from the resulting MLD maps using the median of all available MLD values in a particular grid cell. We selected the median because it is a more representative measure of central tendency for MLD than the mean, as was shown by *de Boyer Montégut et al. (2004)*.

Global phytoplankton C biomass stock was computed from the monthly and overall mission composites that were first downsampled from 9-km to 1^o resolution in log10 space using a 2D convolution kernel of size 12x12. Using composites and downsampling them is a spatio-temporal gap filling technique as it reduces or eliminates data gaps that would bias the global estimate. Since ocean color data are not vertically resolved, we assume that the vertical profile of phytoplankton C biomass is uniform down to the mixed layer depth (MLD). Thus biomass stock was computed by multiplying surface satellite estimates of C biomass by the corresponding MLD value, and all valid pixels were summed after also multiplying by the pixel area. Pixel area was approximated using the area integral on a spherical Earth. Biomass stock was computed for the entire ocean as well the deep ocean, excluding the shelves. For comparison purposes biomass stock was computed also with the B05 and S08 climatologies, as well as the CMIP5 ensemble climatologies. These biomass values were derived using only those pixels (at 1^o resolution) where none of the datasets (MLD, satellite-derived, or CMIP5-based) was missing data. Total areas for which the computation was done were estimated from the 1^o ETOPO1-based bathymetry; total ocean area computed this way ($3.608*10^8$ km^2) is in close agreement with the $3.619*10^8$ km^2 estimate of *Eakins and Sharman (2010)*. The carbon biomass and area calculations exclude the Caspian Sea and other major lake bodies. Overall climatology of NCEP MLD values on a 1-degree grid were also used to calculate global phytoplankton carbon biomass stock as a comparison.

S1.6. Algorithm Validation with In-situ POC Measurements

A validation search was performed against the SeaBASS in-situ bio-optical data set (*Werdell et al., 2003*) maintained by NASA's OBPG (http://seabass.gsfc.nasa.gov/seabasscgi/search.cgi, accessed and match-up search performed Nov. 30, 2015.). In-situ POC measurements from SeaBASS were matched with SeaWiFS $R_{rs}(\lambda)$ determinations at 412, 443, 490, 510 and 555 nm, using SeaWiFS reprocessing R2014.0 (note that the rest of the analysis in this manuscript used reprocessing

R2010.0) and the default match-up criteria provided by the online match-up search tool (link above). All water depths were included. For more details of the match ups, see *Bailey and Werdell (2006)*. The matched-up SeaWiFS $R_{rs}(\lambda)$ values were used to generate the allometric PSD determinations of total phytoplankton carbon biomass, as well as the comparable estimates from the *B05* and *S08* approaches, as done in this manuscript for the satellite imagery (Sect. 2). The empirical correction to the N_o parameter was applied (Sect. 3.7). The match-up analysis then compared the satellite estimates of phytoplankton carbon to the corresponding in-situ POC measurements multiplied by 1/3.

S2. Details on Global Phytoplankton Carbon Stock

Globally integrated mixed-layer algal C biomass values have been previously obtained by integrating remotely sensed Chl vertically and converting it to C using an assumed Chl:C ratio. Such estimates range from 0.30 Gt C to 0.86 Gt C (*Antoine et al., 1996; Behrenfeld and Falkowski, 1997b; Le Quéré et al., 2005*). *Antoine et al. (1996)* provide the highest estimate. They integrated Chl profiles vertically from the surface to whichever was larger between MLD provided by *Levitus (1982)* and the depth where sunlight intensity diminishes to 0.1% of its sea-surface value ($Z_{0.1\%}$). The MLD values of *Levitus (1982)* are likely deeper than those of *de Boyer Montégut et al. (2004)* applied here, because the former are based on considerably larger threshold criteria (0.5°C for temperature and 0.125 kg m^{-3} for density) than the latter (Sect. S1.5). Also, $Z_{0.1\%}$ can exceed MLD in warm oligotrophic waters, which cover a large proportion of the total ocean area. This was the case over ~60% of the global ocean area in the study of *Antoine et al. (1996)*; in these cases they employed non-uniform vertical profiles of Chl (*Morel and Berthon, 1989*). For these reasons, it is expected that the global ocean algal biomass estimate by *Antoine et al. (1996)* will be higher than the values we determined here. Similar reasoning holds for the respective estimates by *Le Quéré et al. (2005)* and *Behrenfeld and Falkowski (1997b)*. Importantly, using alternative MLD fields such as the NCEP-based estimates can cause the estimates of Fig. 2 (also Fig. S7) to almost double (not shown), illustrating that global phytoplankton carbon biomass estimates are sensitive to the choice of integration depths. They are also expected to be sensitive to the assumption of uniformity of vertical profiles.

High latitude areas are observable via ocean color remote sensing for only a part of the year due to low solar angles and the polar night (Fig. S4). As a result, monthly climatological estimates of biomass from the three satellite methods (Fig. S8A) represent less area (between ~85% and ~95% of the total ocean area) which varies with the seasons. The seasonal variation observed can thus be confounded by variation in the observed area. In order to alleviate the problem with varying observable area by SeaWiFS and estimate a more representative global seasonal cycle, areas not observed by SeaWiFS were gap-filled with the corresponding CMIP5 model ensemble data and the monthly global time series were recomputed (Fig. S8B). The main difference between Figs. S8A and S8B is that the seasonal amplitudes of all four data sets are decreased. As a measure of seasonality, we consider the difference between the maximum and minimum values from Fig. S8B, as a percentage of the mean annual signal. From the satellite data sets, the *B05* and *S08* estimates exhibit stronger global seasonality (~28-29%) than our PSD-based approach (~19%). The PSD-based approach exhibits the highest percentage of biomass in the continental shelf areas of all data sets (but see empirical correction in Sect. 3.9 and Fig. S7). The CMIP5 models exhibit significantly stronger seasonality (~85%) that the satellite data

sets. Importantly, the models exhibit a single annual peak in the austral summer, whereas the satellite data sets indicate highest global biomass in the transitional months near the equinoxes. These differences in global seasonality of biomass stock between the satellite data and the models suggest that model representation may need improvement in areas that contribute substantially to the global biomass stock in certain parts of the year, such as the Southern Ocean and/or the North Atlantic. However, satellite data also have issues such as underestimation of Chl in the Southern Ocean (*Dierssen and Smith, 2000; Garcia et al., 2005; Kahru and Mitchell, 2010*), indicating that ocean color products in general may be suspect in this undersampled part of the ocean. The special bio-optical character of the Southern Ocean is evidenced elsewhere (*Uitz et al., 2006*), indicating that regionally tuned satellite algorithms may be required. The area is also hard to observe due to high latitudes and cloudiness. This stresses the need for high quality in-situ observations of this region that contributes significantly to the global biological pump (*Marinov et al., 2008*).

S3. Carbon-based vs. Volume-based PSCs

The carbon-based PSCs constitute a recast of the volume fraction PSCs of *Kostadinov et al. (2010)*. As such, both are PSD-based and are only functions of the PSD slope ξ (the N_o parameter cancels when taking the ratio of Eq. S1 for different size classes), for a given set of allometric coefficients and size limits of integration. This recast to carbon via the allometric relationships leads to a modification of the functions but a preservation of their general shape and tendencies as a function of ξ (Fig. 5). The values of the allometric coefficients are a reflection of the fact that smaller phytoplankton cells are more carbon dense than larger cells (*MDL2000; Moal et al., 1987; Verity et al., 1992*). According to the relationships in MDL2000, the tiniest phytoplankton (ESD = 0.5 μm) contain close to 5.5 times more carbon per unit cell volume than the largest phytoplankton cell considered in this study (ESD = 50 μm) (Fig. S1). This results in higher C-based picoplankton fractions when they are based on C as compared to volume-based ones, for all PSD slopes, and the opposite is true of microplankton (Fig. 5). The sign of this difference depends on the PSD slope for nanoplankton, but for most of the ocean ($\xi > \sim 3.5$) C-based nanoplankton fractions are lower than volume-based values, with the exception of some limited areas with very low PSD values, such as the northern North Atlantic, the confluence zone of the Brazil and Falkland/Malvinas Currents and the thin coastal bands of the Weddell and Ross Seas (not shown).

S4. Details on the C-based PSCs vs. Chl Relationships

For picoplankton, *Hirata et al. (2011)* also observe a sigmoid curve, but it is inflected the opposite way from our Fig. 7B. This could be due to the fact that physiological adaptations may lead to changes in pigment composition without changes in size structure or carbon biomass. A fruitful approach for further investigation would be to focus on blending these two approaches to derive more information about oceanic ecosystems. Both approaches agree remarkably well on the general shape of nanoplankton contribution as a function of eutrophic state (cf. our Figs. 5 and 7C with their Fig. 2b), indicating maximum nanoplankton fraction at transitional, intermediate eutrophic states, with nanoplankton never exceeding about 50%. This is encouraging because this limit of the PSD-based model is a result of the mathematical formulation, whereas the *Hirata et al. (2011)* result is based on empirical diagnostic pigment observations. However, the shape of this curve is

not well observed for high Chl values on the bivariate histogram (Fig. 7C) and *Hirata et al. (2011)* observe a higher maximum than the PSD method. Finally, for microplankton the curve shapes agree well, but *Hirata et al. (2011)* fractions reach values up to 80% for high Chl, whereas our PSD-based algorithm rarely exceeds 50% (cf. our Figs. 5 and 7D with their Fig. 2C).

S5. Sensitivity to PSD Parameters and the Limits of Integration

Here we investigate the sensitivity of the carbon-based products to the input parameters, i.e. the PSD parameters and the limits of integration of Eq. 4. Only the upper limit, D_{max}, is analyzed because there are firm biological reasons to set the lower limit at $D_{min} = 0.5$ μm (Sect. S1.1), while the upper limit is ambiguous (e.g. *Sieburth et al., 1978, Brewin et al., 2010, Uitz et al., 2006, Aiken et al., 2008*). Note, however, that *Hirata et al. (2011)* and *Roy et al. (2013)* use different picoplankton limits. This sensitivity analysis is important because total uncertainties are a function not only of the uncertainties of the inputs, but also of the derivatives of the outputs with respect to the inputs (Eq. S2).

The effect of varying D_{max} from the operational value of 50 μm to 200 μm is largest for low PSD slopes (Fig. S9A) and does not exceed ~25% for fractional C-based nanoplankton, somewhat less for microplankton, and much less for picoplankton. The effect diminished quickly for larger PSD slopes and is quite small for $\xi > 4$ (covering most of the ocean, see histogram in Fig. S9A). Using the operational limit globally may cause an underestimation of microplankton contributions and instead may attribute this carbon mostly to nanoplankton, in the eutrophic productive areas of the ocean, during episodes when cells substantially larger than ~50 μm ESD are present in the bloom. The present algorithm is a proof-of-concept approach that is optimized for global applications, and there are reasons to believe the operational D_{max} choice is the best (Sect. S1.1).

Total phytoplankton carbon concentration is a relatively weak function of the PSD slope (Fig. S9B), especially around $\xi = 4$, where the derivative changes sign. There is less than an order of magnitude variability in carbon over the entire range of PSD slope values. In contrast, total carbon is a very strong function of the N_o parameter (Fig. S9C), leading to ~4 order of magnitude variability with the realistic values for N_o. C is a linear function of N_o (Eq. S1). This strong dependence is a very critical finding, illustrating than total carbon concentrations are driven mostly by N_o since it varies over logarithmic scales; in addition, the uncertainties in N_o are relatively higher and spatially uniform themselves (*KSM09*), accounting for most of the uncertainty in total carbon (Fig. S6A). To first order, efforts to improve carbon retrievals thus need to focus on N_o rather than other sources of error. N_o has a similar effect on the carbon concentration in the different PSCs (Fig. S9C). The effect of varying D_{max} is also shown, indicating that microplankton carbon is the only value affected more significantly, but only within much less than an order of magnitude. In contrast, carbon in the different PSCs is a different function of ξ for each PSC (Fig. S9B), illustrating large variability for microplankton and smaller variability for pico- and nanoplankton. As expected, increasing the PSD slope allocates more carbon to the smaller PSCs (at a fixed N_o). D_{max} variability only affects microplankton and total carbon concentrations at low PSD slopes (Fig. S9B), and this variability is generally smaller than

the quantifiable composite uncertainties (cf. Fig. 8A and Fig. S9B), unlike the effect of D_{max} on fractional PSCs, which can be larger than the quantifiable composite uncertainties for low PSD values (cf. Fig. 8B and Fig. S9A).

S6. Details on Algorithm Assumptions and Additional Sources of Uncertainty

The radiometric ocean color products (i.e. SeaWiFS $R_{rs}(\lambda)$ in this case), which are the initial input for the biomass algorithm, are associated with their own uncertainties, as is the output of the *LAS2006* algorithm. These uncertainties are not easy to quantify on a per-pixel basis and are not provided directly by the algorithm developers. *Loisel et al. (2006)* provide a detailed analysis of error sources for the spectral slope of backscattering. These uncertainties are not included in the error budget presented here. However, efforts are currently underway to provide a reasonable quantification of the effect of those uncertainties on the estimated spectral backscattering and its slope, and thus on the PSD and all downstream products.

The PSD parameters are retrieved from the products of the *LAS2006* algorithm via LUTs, which incorporate certain assumptions and uncertainties as well. A detailed analysis of exogenous sources of uncertainties in the PSD parameters is provided in *KSM09*. Here, a brief summary of the important points is provided. The KSM09 algorithm assumes a power-law PSD as does the calculation of particle volume itself (Eq. 2). While there are some indications that deviations from the power-law can be significant, especially in coastal waters (*Reynolds et al., 2010*), it remains a good first-order approximation especially in global applications (*KSM09* and refs. therein). Furthermore, the applicability of the power-law is assumed to hold over the entire diameter range of optically significant particles, including submicron particles, for which measurements are very scarce. Mie scattering theory (*Mie, 1908*) assumes spherical and homogeneous particles, even though it is clear that these assumptions do not hold for living phytoplankton cells. Violations of these Mie theory assumptions are likely to be more severe in coastal and eutrophic areas where larger cells increase in importance (*KSM09*). To date, the backscattering budget is not satisfactorily closed (e.g. *Stramski et al., 2004*), i.e. there is considerable uncertainty in the sources of backscattering and their relative importance. The assumption of biogenic origin of the backscattering particles is inherited from the interpretation of the PSD-determined volume as biogenic (*Kostadinov et al., 2010*). Note that inorganic particles such as coccoliths and even bubbles can contribute to backscattering (e.g. *Balch et al, 2011*; *Randolph et al., 2014*). The difficulty lies in the complexity and variability of the suspended particle assemblages in natural waters, and the limited theoretical abilities to model scattering (*Quirantes and Bernard, 2004; Clavano et al., 2007*). Recent studies indicate that large phytoplankton may be responsible for more backscattering than Mie theory predicts (*Dall'Olmo et al., 2009*), and that significant fraction of backscattering variance is explained by nanoeukaryotes (*Martinez-Vicente et al., 2013*).

As emphasized already, the expression of the PSCs in relative terms as fractions of C biomass has the distinct advantage of being only a function of the PSD slope ξ (for a given set of the allometric coefficients and limits of integration), since N_o cancels out. For the calculation of the fractional PSCs, the contribution from non-autotrophic POC and non-biogenic particles cancels, as long as their proportions when converted to C are reasonably

constant for a single observation (pixel) across all the size classes. Since N_o is subject to larger uncertainties (*KSM09*) *and* it drives total carbon values to first order (Sect. S5), it is expected that the fractional PSCs are a more reliable and robust product. In contrast, caution should be exercised when interpreting and using absolute carbon values. The main source of uncertainty in N_o is the real part of the index of refraction of the particles, n_p, which is allowed to vary over a wide range in the KSM09 algorithm development. According to Mie theory, fewer particles with a higher real refractive index will cause the same amount of backscattering as would more particles of smaller refractive index and otherwise the same characteristics (e.g. *Wozniak and Stramski, 2004*). This is confirmed in observational data sets (*Neukermans et al., 2012*). Therefore, the wide range of n_p used in the KSM09 LUT construction (1.025 to 1.2) results in large uncertainty in N_o retrievals, which is a measure of particle number concentration. A single LUT is applied globally in the KSM09 algorithm. In the open ocean, for example the oligotrophic gyres, mineral particle influences are expected to be minimal and thus n_p would be closer to 1.05, characteristic of organic particles, rather than closer to 1.2, which is characteristic of mineral particles (e.g. *Wozniak and Stramski, 2004*). Therefore, by assuming larger overall values for n_p the LUT in KSM09 is likely to *underestimate* N_o over the open ocean (by attributing the backscattering to fewer particles of higher n_p than reality), and conversely, to possibly *overestimate* it in coastal areas where mineral particle influence could be more substantial. This may explain the spatial range exaggeration seen in the PSD algorithm's retrieval, as compared to the other satellite approaches or the models (Figs. 1 and S3). The KSM09 algorithm was designed for global operational applications (as is the carbon algorithm presented here), but it is expected that regionalizing the LUT based on *a-priori* knowledge of the specific particle assemblages will improve performance. Importantly, this is the primary direction for improvement of our retrievals of absolute carbon concentrations, as N_o contributes to most of the uncertainty (Fig. S6A). In order to address this issue, an empirical correction to N_o is introduced in Sect. 3.7. Future research should also explore the feasibility of applying the relationship of the real index of refraction to intracellular carbon concentration (*Stramski, 1999*) in conceptually different scattering modelling that uses this relationship to model n_p, rather than treating it as a source of random error as in KSM09. The feasibility of such an approach may improve with the advent of global space-borne hyperspectral ocean color sensors such as PACE.

In contrast to the absolute concentrations, fractional PSC uncertainty is driven predominantly by uncertainties in the allometric coefficients over much of the ocean, and sometimes the PSD slope. Thus, improvements in the fractional PSCs should focus mostly on the allometric coefficients, which come with their own set of assumptions and sources of error, only some of which are quantified as the regression coefficients' confidence intervals in *MDL2000*, i.e. the dispersion of the data around the statistical fit. Sources of such errors could be, for example, combining the data coming from fixed and living cells, autotrophs and heterotrophs, and different morphology (thecate vs. athecate dinoflagellates). Other factors that contributed to the variance of the *MDL2000* data set were (details in *MDL2000*): errors in cell dimension and C content measurements, deviations of cell shapes from the geometric approximations used to compute volume and considerable inter- and intra-specific variability in C_{cell}:V_{cell} ratios. This variability necessitates the use of different allometric relationship for diatoms and other non-diatom large cells (Fig. S1). This choice is based on the recommendation of *MDL2000* that the biomass of mixed plankton be

determined by using one equation for diatoms and another for the remainder of unicellular plankton, treating cells above and below V_{cell} = 3,000 μm^3 differently.

Additional uncertainties that are not captured by the variance of the MDL2000 allometric coefficients also exist. For instance, it is not clear how representative these data are of natural phytoplankton assemblages. The *MDL2000* allometric relationships are based mostly on eukaryotes, with only two data points contributed by cyanophytes (prokaryotic). The bacteria are thus underrepresented in the derivation of the allometric relationships, and they are likely to be important, especially in oligotrophic waters. In addition, the diameter range over which the *MDL2000* relationships were derived was ~1.4 – 200 μm, indicating that we are extrapolating these relationships a bit on the lower end, for submicron particles. Clearly, more laboratory work is required to determine reliably the carbon content of small cells, especially since *Stramski (1999)* observe large uncertainties comparing allometric estimates to their carbon estimates (*Stramski et al., 1995*) for *Synechococcus*.

Growth conditions and growth phase could also significantly affect C_{cell}:V_{cell} ratios (*Davidson et al., 2002*). For example, the dinoflagellate cells that MDL2000 used to derive C_{cell}:V_{cell} relationships were grown in nutrient-replete cultures at a fixed temperature and light-dark cycle, and were harvested during exponential growth phase. However, natural habitats often do not provide ideal conditions that can support continued exponential growth. Mesocosm experiments conducted on a natural plankton community suggest that both nutrient limitation and the proportions of macronutrients may have considerable impacts on cellular C concentrations (*Davidson et al., 2002*). *Moal et al.* (1987) observed a drop in cellular C concentration by between ~10% and ~60% after undergoing a shift from exponential to stationary growth. *Stramski et al. (1995)* observed diel variations in cellular carbon content and intracellular carbon concentration for *Synechococcus* grown under natural light-dark cycles.

Some of the above assumptions are necessary artifacts of the model formulation and clearly have no theoretical basis, such as the application of allometric conversion to non-phytoplankton particles. Others can be improved upon by more detailed knowledge of the ecosystems being studied, e.g. the 1/3 factor and the shape of the PSD. Addressing these assumptions will require more observations and theoretical developments. The algorithm presented here is a first order, proof-of-concept approach meant for global applications. Additional knowledge of the ecosystems being studied can be used to improve the estimates, for example if diatoms are known to be dominating a bloom based on an additional source of information, the allometric relationships specific to diatoms can be applied preferentially instead. Taking a more integrated approach to PFT assessment has been studied (*Raitsos et al., 2008*), and future efforts should explore the possibility to leverage knowledge specific to biomes that are allowed to vary in time and space (*Fay and McKinley, 2014*) to tune the algorithm for them, including the underlying PSD LUTs (see below). Furthermore, dynamic assessment of the POC:living C ratio should become more operationally feasible as more concurrent data become available from the field (*Graff et al., 2012; Graff et al, 2015*).

Supplementary Tables

Table S1. The values of parameters *a* and *b* in the allometric Eq.3, as used in Eq. 4 and Eq. S1 in this study to convert volume to carbon. Coefficient values are from *Menden-Deuer and Lessard (2000)*, their Table 4. Standard deviation of the regression coefficients are given in parentheses. The applicable diameter range for each allometric relationship and the weights applied to it (Eq. S1) are also given. V_{cell} stands for cellular volume.

Coefficient set # (*i*)	Phytoplankton group	*log10(a)* (σ)	*b* (σ)	Diameter range applied to [μm]	Weight w_i (Eq. 5)
1	$V_{cell}^{*} < 3000$ μm^3	-0.583 (0.080)	0.860 (0.030)	0.5 – 17.894	1
2	All except diatoms	-0.665 (0.066)	0.939 (0.021)	17.894-50	0.5
3	Diatoms with $V_{cell} >$ 3000 μm^3	-0.933 (0.226)	0.881 (0.045)	17.894-50	0.5

Table S2. Size limits of integration (cellular diameter in μm) applied to the three allometric relationships in Table S1 for the computation of carbon biomass in each size class. These are the D_{min} and D_{max} values used in Eq. S1, with the corresponding weights w_i (Table S1). The allometric coefficients sets correspond to index *i* in Eqns. S1 and S2. The resulting net volume to carbon relationship used in this study is plotted in Fig. S1.

Size Class/Allometric coefficient Set Index *i*	***i* = 1**	***i* = 2**	***i* = 3**
Picoplankton	0.5 – 2	--	--
Nanoplankton	2 - 17.894	17.894 - 20	17.894 – 20
Microplankton	--	20 – 50	20 – 50
Total C biomass	0.5 – 17.894	17.894 – 50	17.894 – 50

Table S3. Summary of the CMIP5 models that include phytoplankton biomass and primary production. The table includes: spatial resolution in the atmosphere and ocean, list of nutrient tracers, ecology subroutine, phytoplankton functional groups modelled, references, and weight we applied in the inter-model averages.

Model	*Nutrients*	*Ecology module*	*Phytoplankton variables*	*References*	*Weight*
CanESM2	N, (but also accounts for Fe limitation)	NPZD based on Denman and Peña (1999).	Generic phytoplankton	Zahariev et al. (2008)	1
CESM1-BGC	P, N,Fe,Si	MET	Diatoms, small phytoplankton, diazotrophs	Moore et al. (2004), Moore et al. (2006)	1
GFDL-ESM2G (M)	P,N,Fe,Si	TOPAZ2	Large phytoplankton (diatoms, greens, and other large eukaryotes), small phytoplankton (prokaryotic picoplankton and nanoplankton), and diazotrophs	Dunne et al. (2013)	1 (1)
HadGEM2-ES (CC)	N,Fe,Si	Diat-HadOCC (NPZD)	Diatoms, non-diatoms	Palmer and Totterdell (2001)	0.5 (0.5)
IPSL-CM5A-MR (LR)	P,N,Fe,Si	PISCES (from HAMOCC5)	Diatoms, nanophytoplankton (non-diatom). Diatoms differ from nanophytoplankton because they need silicon and more iron and because they have higher half-saturation constants due to their larger mean size.	Aumont and Bopp (2006), Séférian et al. (2013)	0.5 (0.5)
MPI-ESM-MR (LR)	P,N,Fe,Si	HAMOCC5.2 (NPZD)	Generic phytoplankton (Plankton concentration is then subdivided into opal - and calcium carbonate-producing fractions as basis for shell production)	Ilyina et al. (2013)	0.5 (0.5)
MRI-ESM1	P,N	NPZD (Oschiles 2001)	Generic phytoplankton	Yukimoto et al. (2011)	1
NorESM1-ME	P,N,Fe,Si	HAMOCC5.1 (NPZD)	Generic phytoplankton (Plankton concentration is then subdivided into opal - and calcium carbonate-producing fractions as basis for shell production)	Assmann et al. (2010)	1
GISS-E2-H-CC (GISS-E2-R-CC)	N, Fe, Si	NOBM	Diatoms, chlorophytes, cyanobacteria, coccolitophores	Gregg (2008)	1 (1)

Additional References

Aiken, J., N. J. Hardman-Mountford, R. Barlow, J. Fishwick, T. Hirata, and T. Smyth (2008), Functional links between bioenergetics and bio-optical traits of phytoplankton taxonomic groups: an overarching hypothesis with applications for ocean colour remote sensing, *Journal of Plankton Research*, *30*(2), 165-181.

Amante, C. and B.W. Eakins (2009), ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. doi:10.7289/V5C8276M ; last access: Jan. 30, 2015.

Assmann KM, Bentsen M, Segschneider J, Heinze C (2010), An isopycnic ocean carbon cycle model, *Geoscientific Model Development* 3:143-167.

Aumont O, Bopp L (2006) Globalizing results from ocean in situ iron fertilization studies, Global Biogeochemical Cycles 20:15. doi:10.1029/2005gb002591

Bailey, S. W., and P. J. Werdell (2006), A multi-sensor approach for the on-orbit validation of ocean color satellite data products, *Remote Sens. Environ.*, 102, 12– 23, doi:10.1016/j.rse.2006.01.015.

Balch WM, Drapeau DT, Bowler BC, Booth ES, Lyczskowski E, Alley D (2011) The contribution of coccolithophores to the optical and inorganic carbon budgets during the Southern Ocean Gas Experiment: New evidence in support of the "Great Calcite Belt" hypothesis. *Journal of Geophysical Research-Ocean*s Special Issue. VOL. 116, C00F06, doi:10.1029/2011JC006941.

de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004), Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology, *Journal of Geophysical Research-Oceans*, *109*(C12).

Clavano, W. R., Boss, E., and Karp-Boss, L. (2007), Inherent optical properties of non-spherical marine-like particles – from theory to observation, *Oceanography and Marine Biology, An Annual Review*, 45, 1–38.

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.

Davidson, K., E. C. Roberts, and L. C. Gilpin (2002), The relationship between carbon and biovolume in marine microbial mesocosms under different nutrient regimes, *European Journal of Phycology*, *37*(4), 501-507.

Dierssen, H. M., and R. C. Smith (2000), Bio-optical properties and remote sensing ocean color algorithms for Antarctic Peninsula waters, *Journal of Geophysical Research: Oceans*, *105*(C11), 26301-26312.

Eakins, B. W., & Sharman, G. F. (2010). Volumes of the World's Oceans from ETOPO1. *NOAA National Geophysical Data Center, Boulder, CO*.

Garcia, C. A. E., V. M. T. Garcia, and C. R. McClain (2005), Evaluation of SeaWiFS chlorophyll algorithms in the Southwestern Atlantic and Southern Oceans, *Remote Sensing of Environment*, *95*(1), 125-137.

Ilyina T, Six KD, Segschneider J, Maier-Reimer E, Li HM, Nunez-Riboni I (2013) Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, *Journal of Advances in Modeling Earth Systems* 5:287-315. doi:10.1029/2012ms000178

Ingleby, B., and M. Huddleston (2007), Quality control of ocean temperature and salinity profiles - Historical and real-time data, *Journal of Marine Systems*, *65*(1-4), 158-175.

Kahru, M., and B. G. Mitchell (2010), Blending of ocean colour algorithms applied to the Southern Ocean, *Remote Sensing Letters*, *1*(2), 119-124.

Ku, H.H. (1966), Notes on the use of propagation of error formulas, J. Res. Nat. Bur. Stand. Sec. C: Eng. Inst., Vol. 70C, No. 4, p. 263.

Levitus, S. (1982), *Climatological Atlas of the World Ocean*, 190 pp., U.S. Department of Commerce, National Oceanic and Atmospheric Administration, Rockville, MD.

Marinov, I., M. Follows, A. Gnanadesikan, J. L. Sarmiento, and R. D. Slater (2008), How does ocean biology affect atmospheric pCO_2? Theory and models, *Journal of Geophysical Research-Oceans*, *113*(C7).

Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, 114(8), 1791-1804.

Martinez-Vicente, V., G. Dall'Olmo, G. Tarran, E. Boss, and S. Sathyendranath (2013), Optical backscattering is correlated with phytoplankton carbon across the Atlantic Ocean, *Geophys. Res. Lett.*, 40, 1154–1158, doi:10.1002/grl.50252.

Moal, J., V. Martin-Jezequel, R. P. Harris, J. F. Samain, and S. A. Poulet (1987), Interspecific and intraspecific variability of the chemical composition of marine phytoplankton, *Oceanologica Acta*, *10*(3), 339-346.

Moore JK, Doney SC, Lindsay K (2004), Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model, *Global Biogeochemical Cycles* 18. doi:10.1029/2004gb002220.

Moore JK, Doney SC, Lindsay K, Mahowald N, Michaels AF (2006) ,Nitrogen fixation amplifies the ocean biogeochemical response to decadal timescale variations in mineral dust deposition, *Tellus Series B-Chemical and Physical Meteorology*, 58:560-572. doi:10.1111/j.1600-0889.2006.00209.x.

Morel, A., and J. F. Berthon (1989), Surface Pigments, Algal Biomass Profiles, and Potential Production of the Euphotic Layer - Relationships Reinvestigated in View of Remote-Sensing Applications, *Limnology and Oceanography*, *34*(8), 1545-1562.

Milutinović, S., and L. Bertino (2011), Assessment and propagation of uncertainties in input terms through an ocean-colour-based model of primary productivity, *Remote Sensing of Environment*, *115*(8), 1906-1917.

Neukermans Griet , Loisel Hubert , Mériaux Xavier , Astoreca Rosa , McKee David , (2012), In situ variability of mass-specific beam attenuation and backscattering of marine particles with respect to particle size, density, and composition, *Limnology and Oceanography*, 57, doi: 10.4319/lo.2012.57.1.0124.

O'Reilly, J.E., D.A. Siegel, J. Mueller and 22 Coauthors (2000), SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. *NASA Tech. Memo.*, 2000-206892, Vol. 11, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, 49 pp.

Palmer JR, Totterdell IJ (2001) Production and export in a global ocean ecosystem model, *Deep-Sea Res Part I-Oceanogr Res Pap* 48:1169-1198. doi:10.1016/s0967-0637(00)00080-7

Partensky, F., W. R. Hess, and D. Vaulot (1999), Prochlorococcus, a marine photosynthetic prokaryote of global significance, *Microbiology and Molecular Biology Reviews*, *63*(1), 106-127.

Quirantes, A. & Bernard, S. (2004), Light scattering by marine algae: two-layer spherical and nonspherical models, *Journal of Quantitative Spectroscopy and Radiative Transfer* 89, 311–321.

Randolph, K., H.M. Dierssen, M. Twardowski, A. Cifuentes Lorenzen, C.J. Zappa (2014), Optical measurements of small deeply-penetrating bubble populations generated by breaking waves in the Southern Ocean. *J. Geophys. Res. Oceans*, 119, doi:10.1002/2013JC009227.

Seferian R. et al. (2013) Skill assessment of three earth system models with common marine biogeochemistry, *Climate Dynamics* 40:2549-2573. doi:10.1007/s00382-012-1362-8

Stramski, D., A. Shalapyonok, and R. A. Reynolds (1995), Optical characterization of the oceanic unicellular cyanobacterium Synechococcus grown under a day-night cycle in natural irradiance, *J. Geophys. Res.*, 100(C7), 13295–13307, doi:10.1029/95JC00452.

Stramski, D. (1999). Refractive index of planktonic cells as a measure of cellular carbon and chlorophyll a content. *Deep Sea Research Part I: Oceanographic Research Papers*, 46(2), 335-351.

Stramski, D., E. Boss, D. Bogucki, and K. J. Voss (2004), The role of seawater constituents in light backscattering in the ocean, *Prog. Oceanogr.*, 61, 27– 55.

Verity, P. G., C. Y. Robertson, C. R. Tronzo, M. G. Andrews, J. R. Nelson, and M. E. Sieracki (1992), Relationships between cell volume and the carbon and nitrogen content of marine photosynthetic nanoplankton, *Limnology and Oceanography*, *37*(7), 1434-1446.

UNESCO, ICES, SCOR, and IAPSO (1980), *Tenth report of the joint panel on oceanographic tables and standards* 25 pp., UNESCO, Sidney, B.C., Canada.

Wessel, P., and W. H. F. Smith, A (1996), Global Self-consistent, Hierarchical, High-resolution Shoreline Database, J. Geophys. Res., 101, #B4, pp. 8741-8743.

Wieser, M. E., Holden, N., Coplen, T. B., Böhlke, J. K., Berglund, M., Brand, W. A., De Bièvre, P., Gröning, M., Loss, R.D., Meija, J., Hirata, T., Prohaska, T., Schoenberg R., O'Connor, G., Walczyk, T., Yoneda, S. and Zhu, X. K. (2013), Atomic weights of the elements 2011 (IUPAC Technical Report). Pure and Applied Chemistry, 85(5), 1047-1078.

Wozniak, S. B., and D. Stramski (2004), Modeling the optical properties of mineral particles suspended in seawater and their influence on ocean reflectance and chlorophyll estimation from remote sensing algorithms, *Appl. Opt.*, 43, 3489–3503.

Yukimoto et al. (2011) Meteorological Research Institute-Earth System Model Version 1 (MRI-ESM1) – Model Description, *Meteorological Reports of the Meteorological Research Institute*, 64

Zahariev K, Christian JR, Denman KL (2008), Preindustrial, historical, and fertilization simulations using a global ocean carbon model with new parameterizations of iron limitation, calcification, and N-2 fixation. *Progress in Oceanography*, 77:56-82. doi:10.1016/j.pocean.2008.01.007

Supplementary Figures

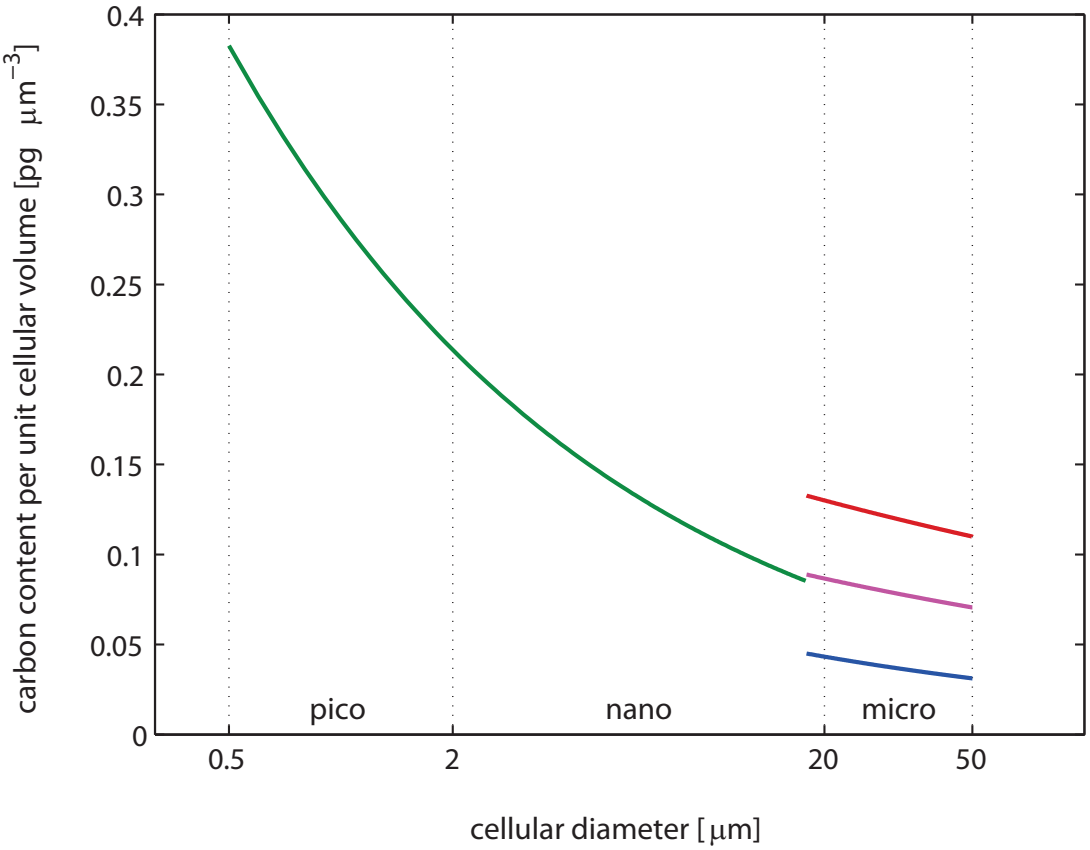


Figure S1. The allometric relationships of *Menden-Deuer and Lessard (2000) (MDL2000)* (their Table 4) as applied for the carbon biomass algorithm (Sects. 2.1, S1.1 and Tables S1 and S2). Carbon content per unit cellular volume is given as a function of cellular diameter. The vertical dotted lines indicate the size ranges of the three phytoplankton size classes (PSCs). The curve exhibits a discontinuity at a diameter of 17.894 μm (V = 3,000 μm^3), because different relationships were reported for phytoplankton below and above that size, respectively. For cells larger than this cutoff diameter, two separate allometric relationships are used (diatoms (blue) and all the rest (red)) and averaged (magenta) for use in the operational algorithm.

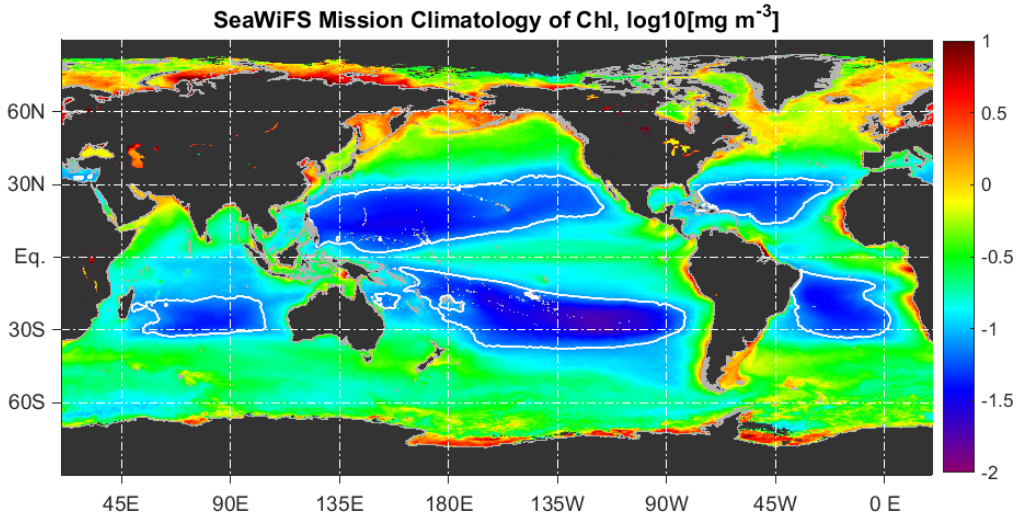


Figure S2. SeaWiFS mission composite (September 1997 – December 2010) of chlorophyll (Chl) concentration (mg m^{-3}), using the standard SeaWiFS algorithm OC4v6.

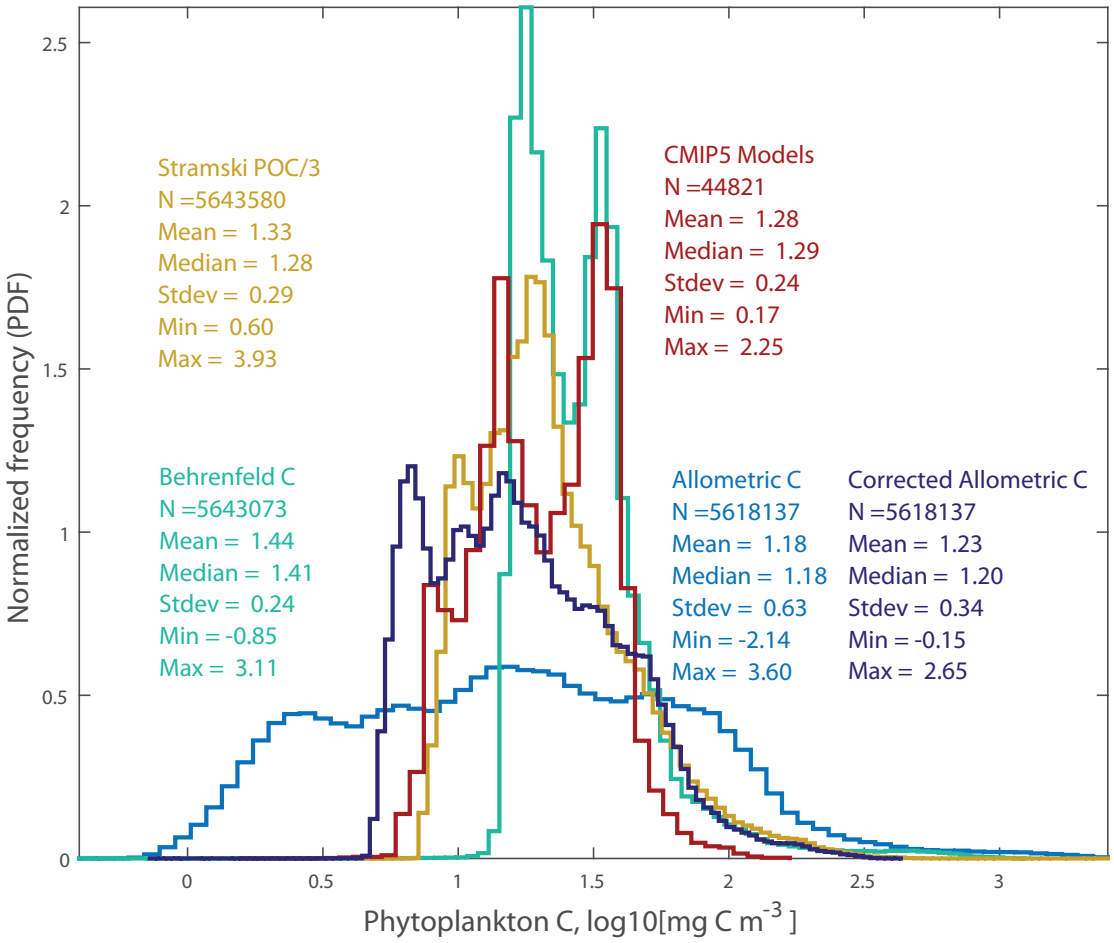


Figure S3. Normalized frequency distributions (probability density functions, PDFs) of the mission mean phytoplankton carbon biomass images of Fig. 1A-D, namely the allometric PSD carbon estimate (light blue), the *Stramksi et al. (2008)* POC retrieval, multiplied by 1/3 (beige), the *Behrenfeld et al. (2005)* method (teal), and the ensemble mean of the CMIP5 models (dark red). The PDF of the allometric PSD estimate of phytoplankton when an empirical correction to the N_o parameter is applied is also shown in darker blue (this is a PDF of the image in Fig. 9, see Sect. 3.7 for details). The x-axis is in log10-space.

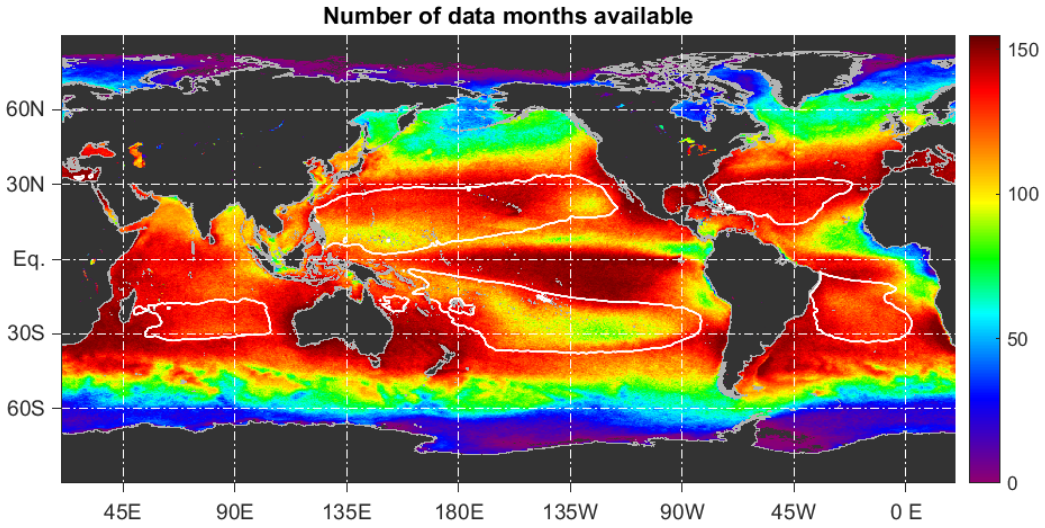


Figure S4. The number of data values contributing to the SeaWiFS mission composite means of the carbon-based products. The number of available monthly data files for the SeaWiFS mission is 157, but the maximum of available data points at any pixel as indicated here is N = 155, reflecting several months with very sparse data in the latest few SeaWIFS years, due to data outages.

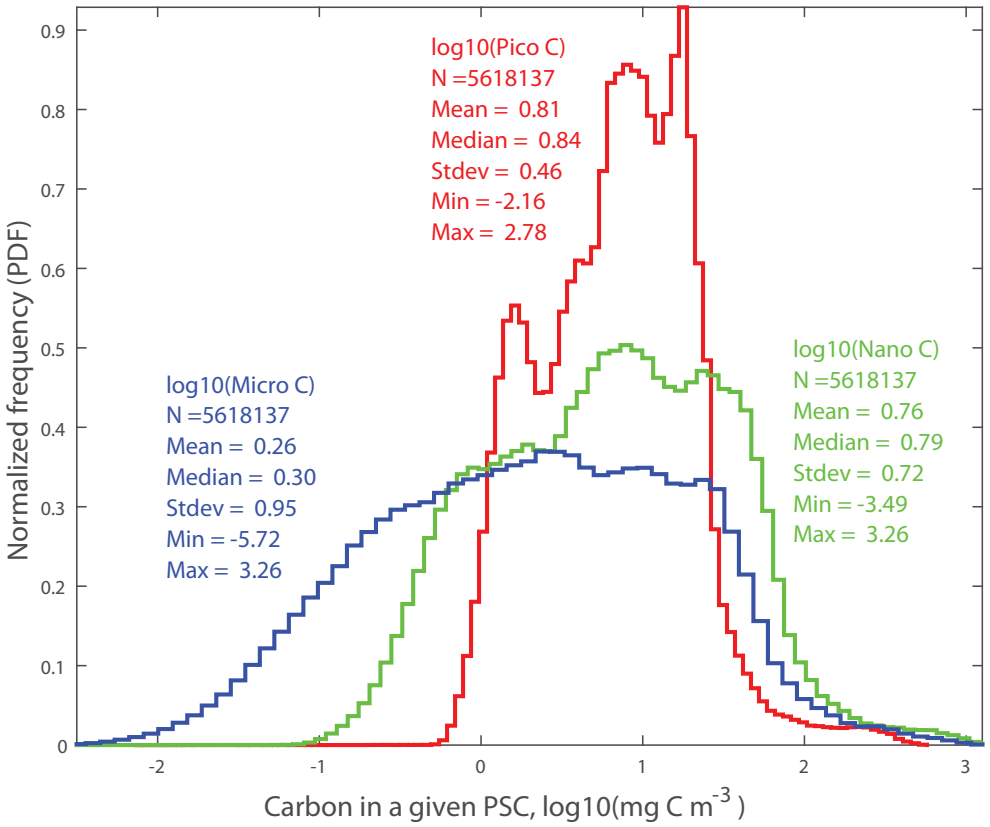


Figure S5. Normalized frequency distributions (probability density functions) computed from the mapped global SeaWiFS mission composites of phytoplankton carbon biomass (in log-10 space) in the three PSCs (Fig. 3A-C) – picoplankton (red), nanoplankton (green) and microplankton (blue).

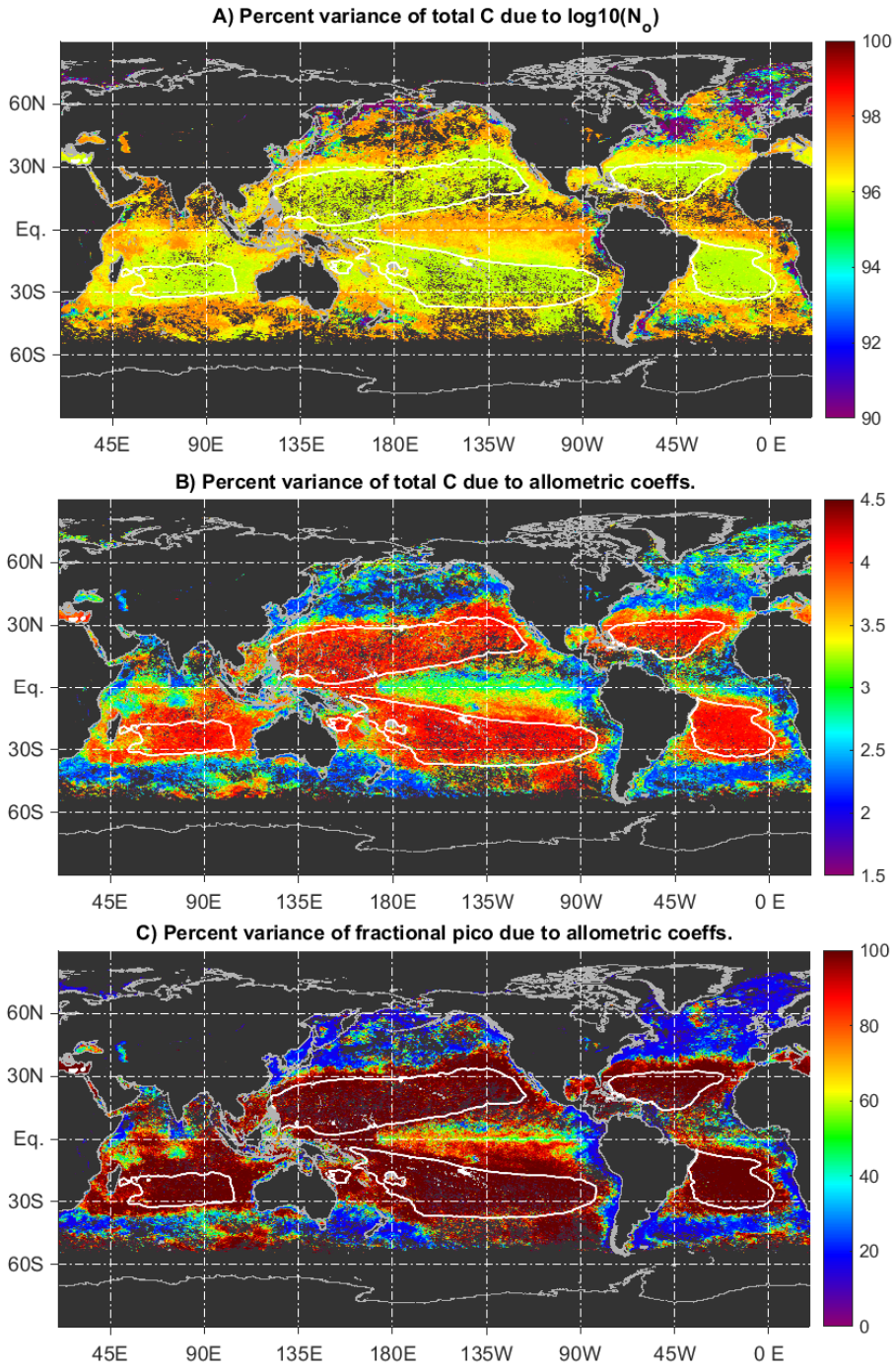


Figure S6. Fraction of uncertainty of total phytoplankton carbon biomass due to A) the N_o PSD parameter, and B) the allometric coefficients. The percent of total variance is shown. The third quantified source of uncertainty, the PSD slope ξ, contributes negligible amounts of variance (<5% for most pixels) and is not shown (the three sources add up to a total of 100% at each pixel). C) The fraction of propagated variance of percent C-based picoplankton due to the allometric coefficients; the remainder to 100% is due to the PSD slope ξ. May of 2004 is shown in all three panels.

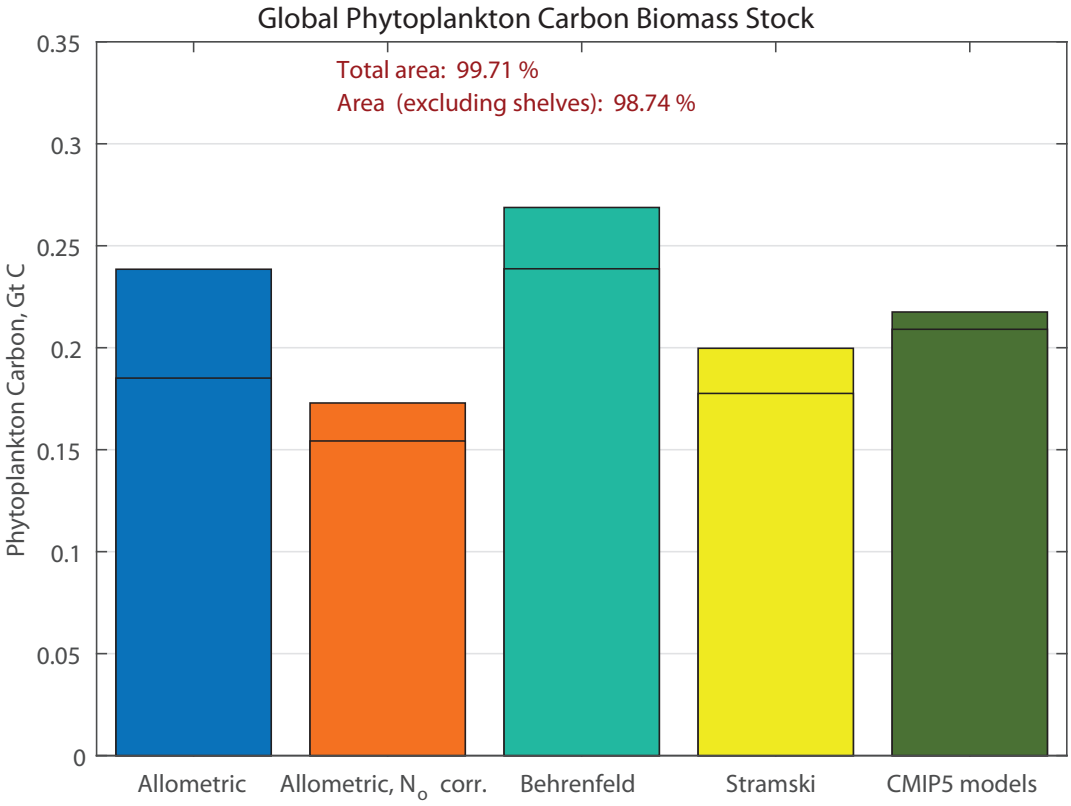


Figure S7. As in Fig.2, but also illustrated are the global phytoplankton carbon stock estimates using the empirically corrected allometric PSD method (orange bars); See Sect. 3.7 for details.

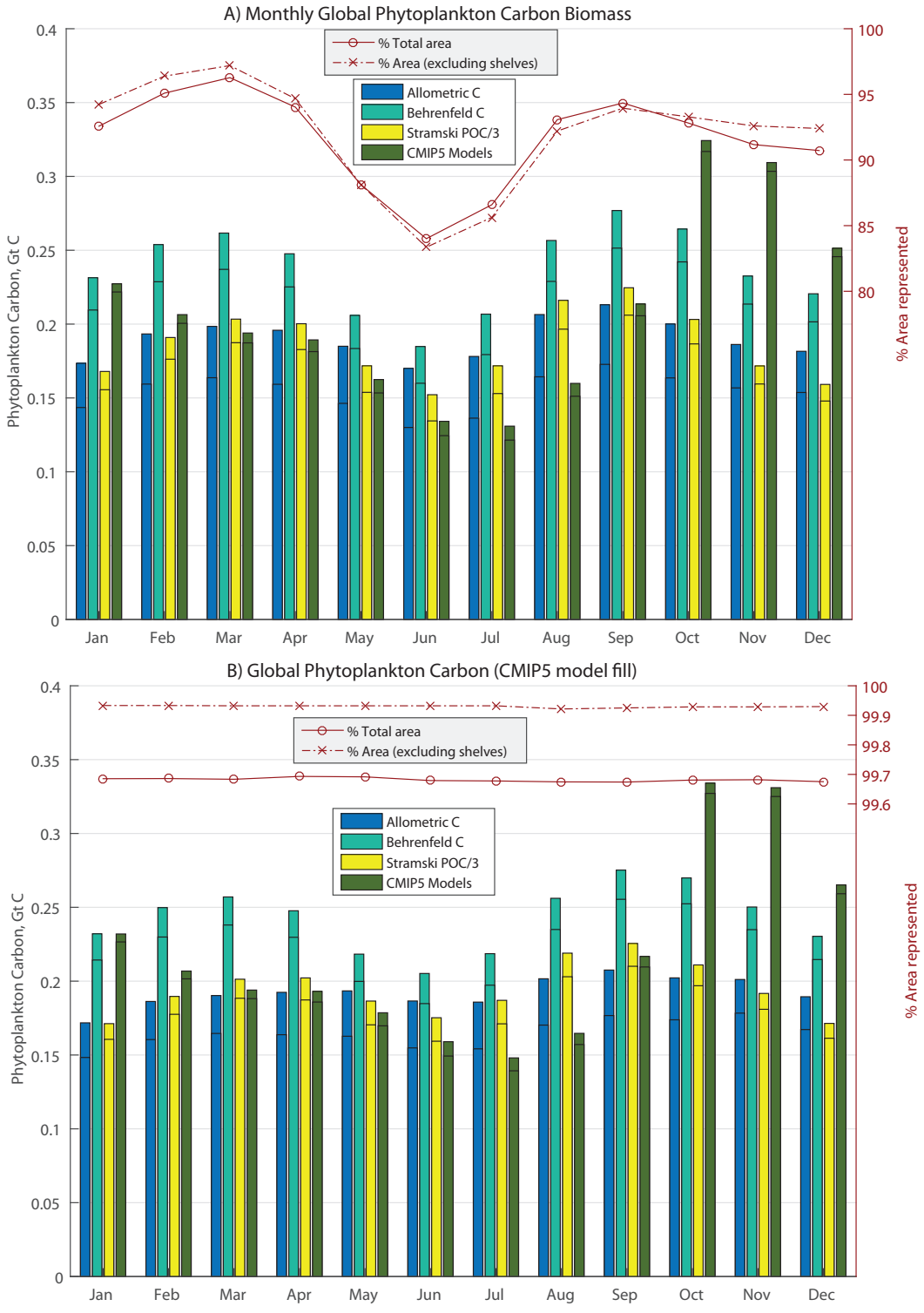
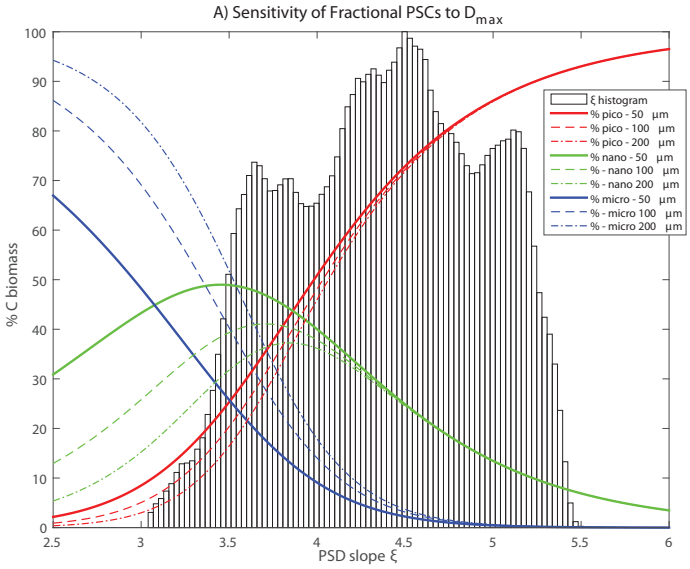
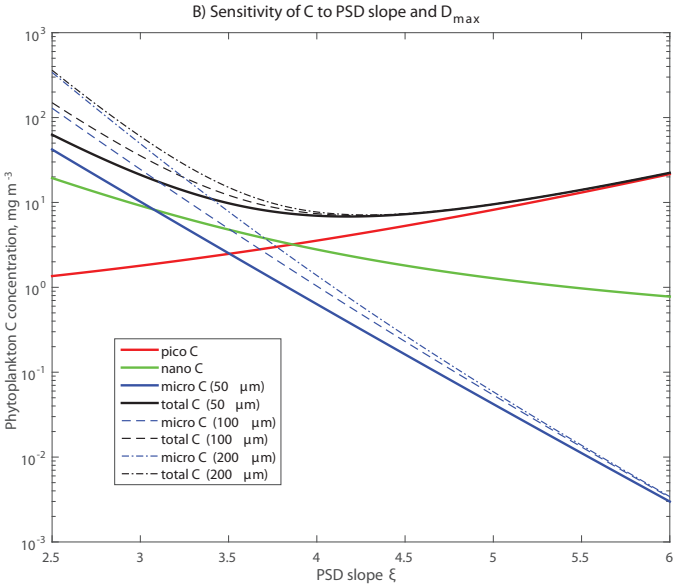


Figure S8. A) Same as in Fig. 2, but for the monthly composite means for the three satellite data sets and the CMIP5 model ensemble mean. B) Same as in A), but with missing SeaWiFS pixels gap-filled with CMIP5 model data in order to represent the entire ocean in the calculation.


A) Sensitivity of Fractional PSCs to D_{max}
% C biomass
PSD slope ξ
ξ histogram
% pico - 50 µm
% pico - 100 µm
% pico - 200 µm
% nano - 50 µm
% - nano 100 µm
% - nano 200 µm
% micro - 50 µm
% - micro 100 µm
% - micro 200 µm


B) Sensitivity of C to PSD slope and D_{max}
Phytoplankton C concentration, mg m^{-3}
PSD slope ξ
pico C
nano C
micro C (50 µm)
total C (50 µm)
micro C (100 µm)
total C (100 µm)
micro C (200 µm)
total C (200 µm)

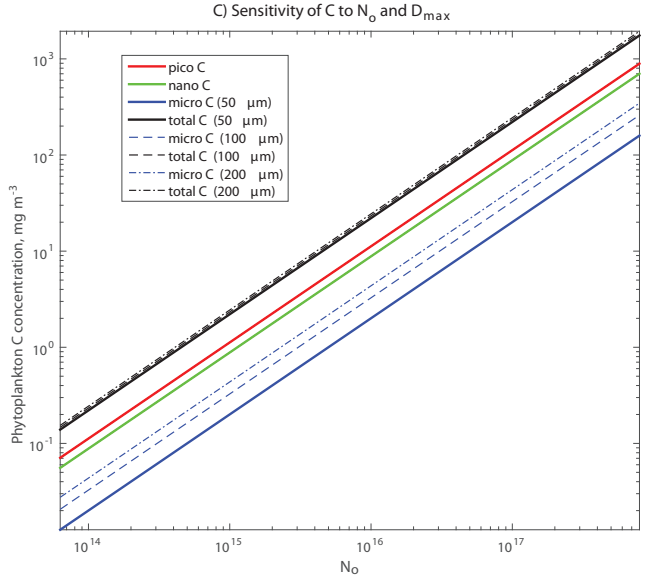

C) Sensitivity of C to N_o and D_{max}
Phytoplankton C concentration, mg m^{-3}
N_o
pico C
nano C
micro C (50 µm)
total C (50 µm)
micro C (100 µm)
total C (100 µm)
micro C (200 µm)
total C (200 µm)

Figure S9. Sensitivity analyses of total and partitioned phytoplankton C biomass to the maximum limit of integration, D_{max} and the PSD parameters ξ and N_o (Eqs. 4 and S1): A) the three PSCs defined as percent contribution to total C biomass, as a function of PSD slope ξ, for three different values of D_{max}, as indicated. The histogram of the SeaWiFS mission composite PSD slope image is included (normalized to the highest count bin); B) Total and partitioned absolute phytoplankton carbon concentration as a function of the PSD slope ξ, when N_o is fixed at 15.5 m^{-4}; C) Total and partitioned absolute phytoplankton carbon concentration as a function of N_o, when the PSD slope is fixed at $\xi = 4$. In panels B) and C), the effect of varying D_{max} on total and microplankton C is also shown, as indicated. The cases corresponding to the operational value (D_{max} = 50 μm) are plotted in bold solid lines in all three panels.