Distributions of phytoplankton carbohydrate, protein and lipid in the world oceans from satellite ocean colour

Article

Accepted Version

Distributions of phytoplankton carbohydrate, protein and lipid in the world oceans from satellite ocean colour

Shovonlal Roy[*1]

[1]Department of Geography and Environmental Science & School of Agriculture, Policy and Development, University of Reading, Whiteknights, Reading RG6 6AB, U.K

January 26, 2018

Conflict of interest: The author declares no conflict of interest.

Running title: Satellite-derived carbohydrate, protein and lipid

[*]Corresponding author. E-mail: shovonlal.roy@reading.ac.uk

Abstract

Energy value of phytoplankton regulates the growth of higher trophic species, affecting the tropic balance and sustainability of marine food webs. Therefore, developing our capability to estimate and monitor, on a global scale, the concentrations of macromolecules that determine phytoplankton energy value, would be invaluable. Reported here are the first estimates of carbohydrate, protein, lipid, and overall energy value of phytoplankton in the world oceans, using ocean-colour data from satellites. The estimates are based on a novel bio-optical method that utilises satellite-derived bio-optical fingerprints of living phytoplankton combined with allometric relationships between phytoplankton cells and cellular macromolecular contents. The annually-averaged phytoplankton energy value, per cubic meter of sub-surface ocean, varied from less than 0.1 kJ in subtropical gyres, to 0.5–1.0 kJ in parts of the equatorial, northern and southern latitudes, and rising to more than 10 kJ in certain coastal and optically complex waters. The annually-averaged global stocks of carbohydrate, protein and lipid were 0.044, 0.17 and 0.108 gigatonnes, respectively, with monthly stocks highest in September and lowest in June, over 1997-2013. The fractional contributions of phytoplankton size classes e.g., picoplankton, nanoplankton and microplankton to surface concentrations and global stocks of macromolecules varied considerably across marine biomes classified as Longhurst provinces. Among these provinces, the highest annually-averaged surface concentrations of carbohydrate, protein, and lipid were in North-East Atlantic Coastal Shelves, whereas, the lowest concentration of carbohydrate or lipid were in North Atlantic Tropical Gyral, and that of protein was in North Pacific Subtropical Gyre West. The regional accuracy of the estimates and their sensitivity to satellite inputs are quantified from the bio-optical model, which show promise for possible operational monitoring of phytoplankton energy value from satellite ocean colour. Adequate *in situ* measurements of macromolecules and improved retrievals of inherent optical properties from high-resolution satellite images, would be required to validate these estimates at local sites, and to further improve their accuracy in the world oceans.

Keywords

Phytoplankton size spectra; ocean colour; carbohydrate; protein; lipid; energy content.

1 Introduction

The autotrophic phytoplankton species in the upper ocean, constituting less than 1% of the entire photosynthetic biomass on the globe, are responsible not only for ∼50% of the global annual carbon-fixation (Falkowski, 2012; Field et al., 1998), but also for providing life-support to marine food-webs through its trophic connections. In addition to their biomass and species composition, the cellular macromolecular contents and energy value of phytoplankton can strongly impact the trophic balance within a marine ecosystem, e.g., by directly impacting the developmental stages of grazers, and influencing the trophic-energy flow affecting the production of higher trophic species (Breteler et al., 2005; Jónasdóttir, 1994; Litzow et al., 2006; Shin et al., 2003). The stoichiometric ratio, i.e., the relative elemental composition of carbon, nitrogen and phosphorous in phytoplankton is known to vary with phytoplankton assemblages across resource gradients in marine biomes (Geider and La Roche, 2002; Martiny et al., 2013). Stoichiometric variations alter the nutritional quality of phytoplankton as food to the grazers (Goldman and Caron, 1985; Sterner and Elser, 2002); and variations in nutrient bound or energy value of phytoplankton affect the stability and oscillatory dynamics of producer-grazer interactions (e.g., Roy et al., 2005; Roy and Chattopadhyay, 2007*a*,*b*). It is, therefore, imperative to monitor the variations in cellular macromolecular contents of marine phytoplankton, on local, regional and global scales. In this context, possibilities of having satellite-based estimates would be invaluable, given that *in situ* observations are often infrequent, and inadequate for monitoring over large spatial scales. Moreover, conducting *in situ* measurements of the macromolecular contents of phytoplankton in the global ocean, would be extremely time consuming and considerably expensive.

Over the last two decades, several satellite-based methods have been developed to extend our capabilities from routinely estimating chlorophyll concentration, to distinguishing phytoplankton functional types (PFTs), in terms of the proportions of chlorophyll either in major taxonomic groups, or in phytoplankton size classes (PSCs) (for more details, see, IOCCG, 2014; Mouw et al., 2017). Some progress has also been made to estimate phytoplankton carbon (Behrenfeld et al., 2005; Kostadinov et al., 2016; Roy et al., 2017; Sathyendranath et al., 2009), and carbon-based classification of PSCs (Kostadinov et al., 2016; Roy et al., 2017), from ocean colour. However, strong variations in phytoplankton cellular carbon and carbon-based macromolecules, with taxa, cell morphology, and environmental conditions such as ambient light and available nutrient (Hitchcock, 1982; Marañón, 2008; Marañón et al., 2013; Menden-Deuer and Lessard, 2000; Strathmann, 1967), impose additional layers of difficulties in converting the satellite-derived estimates of chlorophyll or carbon-to-macromolecular concentrations. Certain phytoplankton macromolecules, such as cellular fatty acids, strongly vary (e.g., between 1% and 85%, Chisti, 2007), not only among algal groups and species, but also within a specific algal group, e.g., diatoms under different culture conditions (Opute, 1974). In addition to laboratory cultures, essential fatty acids in phytoplankton have also been reported to vary with oceanographic conditions, such as sea-surface temperature and chlorophyll-a, on regional scales (e.g., Budge et al., 2014; Pethybridge et al., 2015). However, progress is yet to be made to estimate the variations in total phytoplankton lipid from satellite

data, on a global scale. Moreover, no method exists yet to estimate from satellite, either on a regional or global scale, the spatiotemporal variations of other essential phytoplankton macromolecules, such as carbohydrate or protein. Given that the proportional contributions of these macromolecules determine the energy value of phytoplankton, it would be useful to develop an advanced ocean-colour-based method for estimating the macromolecular concentrations in the ocean waters.

In this paper, the cellular macromolecular contents of marine phytoplankton, in particular, the concentrations of carbohydrate, protein and lipid are estimated on a global scale, for the first time, based on ocean-colour data from satellite remote-sensing. To do so, a novel method is derived that utilises light-absorption coefficients of phytoplankton (a_{ph}) - an inherent optical property (IOP) retrievable from ocean colour (e.g., IOCCG, 2006), coupled with allometric relationships between phytoplankton cells and their cellular macromolecular contents, reported in the literature (Hitchcock, 1982; Marañón, 2008; Marañón et al., 2013, 2007; Menden-Deuer and Lessard, 2000; Moal et al., 1987; Peters, 1983; Strathmann, 1967). The method builds on a semi-analytical algorithm for retrieving the exponent of the phytoplankton size spectrum from satellite ocean colour, developed recently by Roy et al. (2013, 2011). The concentrations of the total macromolecular contents are further partitioned according to their contributions in three bulk PSCs, namely, picoplankton, nanoplankton and microplankton. The estimates are obtained over the global ocean, and for different marine biomes represented by Longhurst oceanographic provinces (Longhurst, 1995, 1998). Further, insights on the estimation uncertainties are provided through detailed sensitivity analyses, highlighting the possibilities of further improvements of the estimates, with the expectation that the input satellite data would further improve, as the satellite era enters into higher temporal and spatial resolution.

2 Methodology

2.1 Satellite validation

Global 4-km, level-3 mapped chlorophyll concentration, remote-sensing reflectance and the IOPs were obtained from the European Space Agency's Ocean Colour Climate Change Initiative (OC-CCI) project (freely available on `http://www.esa-oceancolour-cci.org`). The OC-CCI data were produced by merging ocean-colour data from three satellite sensors: NASA-SeaWiFS, NASA-MODIS-Aqua and ESA-MERIS; further details on OC-CCI, including data processing, temporal consistency of the data products and details of the algorithms used, can be found in Brewin et al. (2015); Müller et al. (2015). Monthly climatologies of the mixed-layer depth were obtained on $0.5^\circ \times 0.5^\circ$ spatial grid from Monthly Isopycnal & Mixed-layer Ocean Climatology (MIMOC, Schmidtko et al., 2013, available freely on `http://www.pmel.noaa.gov/mimoc/`). To obtain depth-integrated estimates of the satellite-derived products from OC-CCI, the mixed-layer depths were remapped onto OC-CCI 4-km grids using nearest-neighbour interpolation by implementing MATLAB2015b interpolation

routine (similar to previous studies, e.g., Roy et al., 2017).

A sufficiently large global *in situ* dataset on phytoplankton carbohydrate, protein and lipid that would ideally be required to validate the satellite-based estimates was unavailable. The historical *in situ* measurements on carbohydrate, protein and lipid, which were already compiled by Finkel et al. (2016*a*), did not cover the period over which satellite data (e.g., OC-CCI v2) were available (i.e., September 1997 onwards). These constraints on hindered satellite validation exercise in different oceanographic conditions.

Whilst direct measurements on carbohydrate, protein and lipid were unavailable, large datasets on *in situ* phytoplankton abundance were available, e.g., those compiled in marine biodiversity database (Sal et al., 2013), which included phytoplankton cell counts from samples collected in different oceanographic cruises between 1992 and 2002, partly covering the satellite period. Owing to the constraints on direct measurements, a validation exercise was attempted by converting the *in situ* data on phytoplankton abundance (Sal et al., 2013) into estimates of phytoplankton macromolecular concentrations, using allometric relationships from the literature (Finkel et al., 2016*a*). To do so, a subset of phytoplankton abundance data (Sal et al., 2013) that overlapped with the OC-CCI v2 temporal coverage (September 1997 - December 2013) were considered, and the concentrations of phytoplankton carbohydrate, protein and lipid were computed using the information on phytoplankton cell size (reported in Sal et al., 2013) and the corresponding allometric relationships (reported in Finkel et al., 2016*a*). This subset included 250 samples collected from 1997 to 2002, across various oceanographic regions (see Section 3.2, for the geographic locations); and consisted of 943 species of diatom, dinoflagellate and coccolithophores with equivalent-spherical diameter ranging from 1.34 μm to 50 μm (to be consistent with the size range of microplankton assumed within the algorithm, only the species with diameter <50 μm were considered). This cell-diameter range covered nanoplankton, microplankton and a part of picoplankton. To be consistent with previous studies (Roy et al., 2013, 2017), the diameter ranges of the three phytoplankton size classes used in the model were picoplankton: 0.2–2 μm, nanoplankton: 2–20 μm, and microplankton: 20–50 μm. Satellite matched-up chlorophyll concentrations and IOPs were retrieved from OC-CCI data archive. Given that the sampling times were mostly within the early years of SeaWiFS coverage (and SeaWiFS was the only contributing ocean-colour sensor over 1997-2002), a large number of gaps in satellite data were identified. To maximise the number of validation data points, match-ups from composite satellite images on daily ($n = 39$) and monthly ($n = 249$) scales were used.

The global annual stocks of phytoplankton carbohydrate, protein and lipid within the oceanic mixed layer were computed from the estimated surface concentrations, grid-by-grid, using the available mixed-layer depth values obtained from MIMOC (no specific depth profiles of the macromolecular concentrations were known from either *in situ* or remote sensing, on a global scale).

2.2 Relating the size spectrum of phytoplankton to its cellular macromolecular concentrations

Studies have shown that phytoplankton cell size strongly determines its cellular concentrations of chlorophyll, carbon and carbon-based macromolecules through allometric relationships (Hitchcock, 1982; Marañón, 2008; Marañón et al., 2013, 2007; Menden-Deuer and Lessard, 2000; Moal et al., 1987; Peters, 1983; Strathmann, 1967). The allometric relationship between the cellular concentration of a macromolecule ($[M]_{cell}$, expressed in the units of pg cell^{-1}) and the volume of a phytoplankton cell (V_{cell}, in μm^3) can be described by the canonical equation: $[M]_{cell} = a_M\, V_{cell}^{b_M}$; where, M stands for the macromolecule that can be carbohydrate, protein or lipid, and a_M, b_M are the allometric parameters with magnitudes specific to a macromolecule M. For a given macromolecule, a_M and b_M would remain constant across the size spectrum of phytoplankton cells. Assuming that the particle size distribution of phytoplankton cells follows the power law (McCave, 1984; Reynolds et al., 2010; Sheldon et al., 1972), the number of phytoplankton cells with equivalent spherical diameter D per unit volume of seawater can expressed as: $N(D) = k\, D^{-\xi}$, with ξ as the exponent of the phytoplankton size spectrum, and k as a constant related to the abundance of the total population. Following Roy et al. (2013), the concentration of phytoplankton chlorophyll-a (B_{total}, mg Chl m^{-3}) within the cell-diameter range [D_{min}, D_{max}] can be expressed as a product of the number of phytoplankton cells within that size class, the volume of each cell ($\pi\, D^3/6$), and the intracellular concentration of chlorophyll-a c_i (in mg m^{-3}, parameterised as $c_i = c_0\, D^{-m}$ with the magnitudes of $c_0 = 3.9 \times 10^6$, and $m = 0.06$ by Roy et al., 2011 using the *in situ* measurements of Marañón et al., 2007), as follows:

$$B_{total} = \int_{D_{min}}^{D_{max}} \left[\left(\frac{\pi}{6}\, D^3\right)\, (c_0\, D^{-m})\, \left(kD^{-\xi}\right)\right]\, dD = \left(\frac{\pi}{6}\, k\, c_0\right)\, \frac{D_{max}^{4-\xi-m} - D_{min}^{4-\xi-m}}{4-\xi-m}. \quad (1)$$

Similarly, the total concentration of the macromolecule M (in mg m^{-3}) due to all phytoplankton cells within a diameter range [D_{min}, D_{max}] can be expressed as a product of the number of cells and the cellular concentration $[M]_{cell}$:

$$\begin{aligned} [M]_{total} &= \int_{D_{min}}^{D_{max}} [N(D) \times [M]_{cell}]\, dD = \int_{D_{min}}^{D_{max}} \left(kD^{-\xi}\right) \left[10^{-9}\, a_M\, \left(10^{18}\, \frac{\pi}{6}\, D^3\right)^{b_M}\right]\, dD, \\ &= 10^{-9}\, k\, a_M\, \left(10^{18}\, \frac{\pi}{6}\right)^{b_M}\, \left(\frac{D_{max}^{3b_M-\xi+1} - D_{min}^{3b_M-\xi+1}}{3b_M-\xi+1}\right); \end{aligned} \quad (2)$$

with the condition that $[M]_{total} \rightarrow \left[10^{-9}\, k\, a_M\, (10^{18}\, \pi/6)^{b_M}\, \log_e\left(D_{max}/D_{min}\right)\right]$, when $\xi \rightarrow (3b_M + 1)$, applied to avoid division by zero. The factors 10^{-9} and 10^{18} are associated with the conversions of units from pg to mg, and m^3 to μm^3 respectively. Using Eqs. (1) and (2), the ratio of the macromolecular concentration to the chlorophyll concentration (χ_M) can be expressed as:

$$\chi_M = \frac{[M]_{total}}{B_{total}} = \frac{10^{-9}\, a_M\, (10^{18}\, \pi/6)^{b_M}}{(\pi/6)\, c_0}\, \left(\frac{D_{max}^{3b_M-\xi+1} - D_{min}^{3b_M-\xi+1}}{D_{max}^{4-\xi-m} - D_{min}^{4-\xi-m}}\right)\, \left(\frac{4-\xi-m}{3b_M-\xi+1}\right). \quad (3)$$

Note that the expression of macromolecule-to-chlorophyll ratio χ_M in Eq. (3) does not depend on the parameter k appearing in Eqs. (1) and (2). So, once χ_M is computed, M_{total} can be computed from the observed value of B_{total} as:

$$M_{total} = \chi_M \, B_{total}, \tag{4}$$

provided that ξ, a_M and b_M of the population are known (see Sections 2.3, 2.4).

2.3 Size-partitioned cellular contents of phytoplankton

Assuming that the total biomass of phytoplankton is a sum of the biomasses of n non-overlapping PSCs defined by cell-diameter ranges $[D_i,\, D_j]$ with $0 \leq i < j \leq n$, $[M]_{total} = \sum [M]_{ij}$, where $[M]_{ij}$ denote the macromolecular concentration within the size class $[i, j]$. It follows from Eq. (4), that $[M]_{ij} = \chi_{M\,ij}\, B_{ij}$, with $\chi_{M\,ij}$ and B_{ij}, respectively, are the macromolecule-to-chlorophyll ratio and the concentration of chlorophyll B_{ij} in the size class $[D_i,\, D_j]$, where $\chi_{M\,ij}$ follows directly from using Eq. (3):

$$\chi_{M\,ij} = \frac{10^{-9}\, a_M \, (10^{18}\,\pi/6)^{b_M}}{(\pi/6)\, c_0} \left[\frac{D_j^{3b_M-\xi+1} - D_i^{3b_M-\xi+1}}{D_j^{4-\xi-m} - D_i^{4-\xi-m}}\right] \left[\frac{4-\xi-m}{3b_M-\xi+1}\right], \tag{5}$$

and the expression of B_{ij} is taken from Roy et al. (2013), so that,

$$[M]_{ij} = \chi_{M\,ij}\, B_{ij} = \chi_{M\,ij} \left(\frac{D_j^{4-\xi-m} - D_i^{4-\xi-m}}{D_{max}^{4-\xi-m} - D_{min}^{4-\xi-m}}\right) B_{total}; \tag{6}$$

and therefore,

$$[M]_{total} \;=\; \sum_{i=0,\, j=i+1}^{i=n-1,\, j=n} [M]_{ij} = \frac{B_{total}}{D_{max}^{4-\xi-m} - D_{min}^{4-\xi-m}} \sum_{i=0,\, j=i+1}^{i=n-1,\, j=n} \left[\chi_{M\,ij} \left(D_j^{4-\xi-m} - D_i^{4-\xi-m}\right)\right] \tag{7}$$

Also, the fraction of $[M]_{ij}$ to $[M]_{total}$ can be computed as:

$$F_{M,ij} = \frac{[M]_{ij}}{[M]_{total}} = \frac{\chi_{M\,ij}, \left(D_j^{4-\xi-m} - D_i^{4-\xi-m}\right)}{\sum_{i=0,\, j=i+1}^{i=n-1,\, j=n} \left[\chi_{M\,ij} \left(D_j^{4-\xi-m} - D_i^{4-\xi-m}\right)\right]}. \tag{8}$$

Using the equations derived above, the concentrations of carbohydrate, protein and lipid can be partitioned into any number of PSCs. However, for the sake of discussion, in this study, the estimates are obtained for three major PSCs, namely, picoplankton, nanoplankton and microplankton, with cell-diameter bounds $[D_0,\, D_1]$, $[D_1,\, D_2]$ and $[D_2,\, D_3]$, respectively, where $D_0 = 0.25$ μm, $D_1 = 2$ μm, $D_2 = 20$ μm, and $D_3 = 50$ μm based on previous studies (Roy et al., 2013; Sieburth et al., 1978; Vidussi et al., 2001).

2.4 Allometric parameters a_M and b_M from the literature, and retrieval of ξ from satellite data

The allometric parameters a_M and b_M corresponding to phytoplankton species are reported in several studies e.g., Finkel et al. (2016*a*); Hitchcock (1982); Menden-Deuer and Lessard (2000); Moal et al. (1987). More recently, Finkel et al. (2016*a*) compiled a large database of macromolecular concentrations in various eukaryotic microalgae from 53 published studies, covering various taxonomic groups, culture conditions and growth phases; and reported the allometric relationships between cell volume and concentrations of carbohydrate, protein, and lipid in phytoplankton. In the current study, a_M and b_M are fixed based on Finkel et al. (2016*a*) (see, their Table-II), and their reported values along with the confidence intervals are used for estimating the macromolecular concentrations and performing uncertainty analyses described in Section 2.5).

The exponent of the phytoplankton size spectrum ξ is retrieved from the specific-absorption coefficient of phytoplankton at 676 nm using a semi-analytical ocean-colour algorithm developed by Roy et al. (2013). For completeness, the major steps of this methodology are described in the Supplementary Materials, without fully reproducing it from Roy et al. (2013). However, for further details on the parameterisation and optimization steps related the retrieval of ξ, readers are referred to Roy et al. (2013, 2011).

2.5 Uncertainties and biases

Although the method described above is founded on theories of light-absorption properties and cellular allometric relationships of phytoplankton, the estimates need to be validated against direct *in situ* measurements, which are currently unavailable. This limitation raises the possibility of bias and uncertainties in satellite products at each pixel, leading to biased estimates of the macromolecules on a global scale. The inaccuracy of the estimates may arise from several sources, the most prominent of which is the uncertainties associated with the satellite products used as inputs to the model, e.g., chlorophyll-a and absorption coefficients of phytoplankton. The uncertainties in chlorophyll-a retrievals for optically complex (Case II) waters are considerably large, when compared within those for the open oceans (Case I waters), mainly due to the limitations of the empirical chlorophyll algorithms used (e.g., IOCCG, 2000). The absorption coefficients of phytoplankton, on the other hand, being an IOP are retrieved generally by semi-analytical algorithms, the performance of which also vary for optically complex waters (e.g., IOCCG, 2006).

In the coastal oceans and optically complex waters, the retrievals are affected due to the presence of high concentration of coloured-dissolved organic matters (CDOM), sediments, other suspended materials and water constituents that interfere with light penetration and reflectance (IOCCG, 2000). Uncertainties in remote sensing retrievals can further be attributed to clouds, ice covers, solar zenith angles, sun glint, atmospheric dusts and aerosols (e.g., IOCCG, 2000; Maritorena et al., 2010). Thus, the satellite-derived estimates of carbohydrate

protein and lipid presented on global maps (in the result section) comes with uncertainty and bias, an accurate estimation of which would be possible only when adequate *in situ* measurements on these quantities become available.

Nevertheless, to understand and quantify the overall uncertainty levels in the satellite-derived estimates, a model sensitivity analysis was carried out. Theoretically, accurate estimations of the macromolecular concentrations in phytoplankton based on the above method would depend on the allometric parameters (a_M and b_M) and the estimates of ξ. The retrieval of ξ further depends on satellite-derived estimates of chlorophyll-a and a_{ph}. Using Eqs. (1-3), the relative sensitivities of the estimates of M_{total}, i.e., $\frac{\Delta M_{total}}{M_{total}}$, can be computed as a combined function of $\frac{\Delta \xi}{\xi}$, $\frac{\Delta a_M}{a_M}$, and $\frac{\Delta b_M}{b_M}$. Following Roy et al. (2013), where $\frac{\Delta \xi}{\xi}$ are reported pixel-by-pixel in the global ocean, a maximum overall $\frac{\Delta \xi}{\xi}$ in the range 0–25% is considered. For $\frac{\Delta a_M}{a_M}$ and $\frac{\Delta b_M}{b_M}$, the half of the 95% spread with respect to the mean levels reported by Finkel et al. (2016*a*) are considered. The resultant $\frac{\Delta M_{total}}{M_{total}}$ are then computed pixel-by-pixel, as percentages of the default estimates. So, without the availability of adequate *in situ* measurements, the uncertainties discussed in the following sections should be interpreted as model-based uncertainties, and not as those based on the *in situ* observations.

3 Results and discussion

3.1 Macromolecular concentrations across phytoplankton size range

The ratios of carbohydrate-to-chlorophyll (χ_{carbo}), protein-to-chlorophyll (χ_{prot}) and lipid-to-chlorophyll (χ_{lipid}) increase with ξ within the ranges given by [5.0, 9.5], [7.1, 48.9] and [3.1, 32], respectively (Fig. 1a). For any given value of ξ, χ_{prot} is higher than χ_{carbo} and χ_{lipid}. For low values of ξ, χ_{lipid} is lower than χ_{carbo}, but it increases more rapidly with the assemblages of small phytoplankton cells, and so, for high values of ξ, χ_{lipid} is significantly higher than χ_{carbo} (Fig. 1a).

The proportions of carbohydrate, protein and lipid increase with ξ in picoplankton (Fig. 1b), and decrease with ξ in microplankton (Fig. 1d), but are unimodal in nanoplankton having magnitudes typically less than 50% with highest values in the middle rage of ξ (Fig. 1c). At any given level of ξ, the proportion of lipid in picoplankton is higher than that of carbohydrate or protein (with carbohydrate < protein < lipid) (Fig. 1b); but in microplankton the order is reversed to carbohydrate > protein > lipid (Fig. 1d). For nanoplankton these proportions alter from carbohydrate < protein < lipid at the lower end of ξ to carbohydrate > protein > lipid at the higher end of ξ (Fig. 1c). These results show strong dependencies of phytoplankton size structure on the available macromolecular concentrations with implications on their stocks in mixed populations of phytoplankton.

For carbohydrate estimates, the relative uncertainties would be <30% for $3.25 < \xi < 5$ (typically representing small-cell dominated populations), but would increase up to 60% at the lower end of ξ (typically representing large-cell dominated populations) (Fig. 1e, Table 1).

For protein estimates (Fig. 1f), the relative uncertainties would be <40% across the range of ξ provided that the relative uncertainty in ξ is <10%. If the relative uncertainties in ξ are >15%, the uncertainties in protein would increase to >60% typically for $3.25 < \xi < 4.5$, but would generally remain within <40% for populations dominated by either large or small cells (i.e., at the low and high ends of ξ, see Table 1 for more details). For lipid estimates, the relative uncertainties would be similar to those for protein: <40% for the low and high ends of ξ, but >60% for the mid-range of ξ, if the uncertainties in ξ is >15% (Fig. 1g). Further details on these uncertainty estimates for various combinations of uncertainties in ξ estimates (based on Fig. 1e-g) are summarised in Table 1, and the propagations of the uncertainties in the global ocean are discussed in Sections 3.7.

3.2 Comparison with estimates based on *in situ* abundance data

The matched-up *in situ* data were from specific cruises (see, Fig. 2a) with moderate sample size having non-normal distribution; therefore, non-parametric statistics were implemented, in particular, Spearman's correlation instead of Pearson's, and other non-parametric matrices following Werdell et al. (2009). The *in situ* and satellite-based estimates generally follow the 1 : 1 line, but with some level of spread around it (Fig. 2b-d, Supplementary Fig. S1), with significant correlations (Spearman's ρ) between them on linear scale, for carbohydrate: $\rho = 0.25$, $p < 0.001$; protein: $\rho = 0.24$, $p < 0.001$; and lipid: $\rho = 0.23$, $p < 0.001$ (Fig. 2b-d). The root mean squared error (RMSE) and bias of the estimates vary for carbohydrate (RMSE 10.20, bias $-7.28\,\mathrm{mg\,m^{-3}}$), protein (RMSE 21.55, bias $-10.93\,\mathrm{mg\,m^{-3}}$) and lipid (RMSE 9.77, bias $-4.87\,\mathrm{mg\,m^{-3}}$). As expected, the RMSE and bias for daily match-ups, turn out to be lower than those for monthly match-ups (see, Supplementary Table S1); but in both cases their magnitudes are within a reasonable range, when compared with those for other derived products, such as phytoplankton carbon (Kostadinov et al., 2016; Roy et al., 2017).

Following Werdell et al. (2009), three further metrices are computed for comparing the estimates with daily (monthly) match-ups: the median satellite-to-in-situ-ratio (median ratio, found to be 0.51(0.71), 0.59 (0.73), and 0.59 (0.73) respectively), the median of the relative-percent difference (median RPD, found to be −49.41 (−29.36), −40.86 (−27.38) and −41.11 (−26.75), respectively), and the semi-interquartile percent differences (SIQ-PD, found to be −48.50 (−67.65), −50.66 (−65.08) and −51.36 (−63.82), respectively) (see, Supplementary Table S1). The median RPDs and SIQ-PDs are lowest for lipid estimates, followed by those for protein and carbohydrate (Supplementary Table S1). The median ratios are < 1, suggesting that the algorithm would generally underestimate the macromolecular concentrations (Fig. 2e). Also, the algorithm seems to produce relatively less natural variability of the macromolecular concentrations, in comparison with those estimated from the *in situ* abundance data (Fig. 2e). However, it is worth mentioning that the median ratio, median RDP, SIR-PD for SeaWiFS chlorophyll were reported (Werdell et al., 2009) to be in the ranges [1.7, 81.5], [−34.7, 122.3], and [0.88, 1.69], respectively. Therefore, in terms of these metrices, the accuracy of the current estimates of the macromolecular concentrations are generally comparable with that reported for SeaWiFS chlorophyll.

Nevertheless, these comparisons would be affected by several layers of uncertainties associated with the *in situ* and satellite estimates. For example, prominent natural variability of cell size of the 943 phytoplankton species would alter the *in situ* estimates of carbohydrate, protein and lipid, which were not possible to include in the *in situ* calculations; and the uncertainties in satellite inputs (chlorophyll, IOPs) would also affect the satellite retrievals of ξ (also see, Section 3.7).

3.3 Phytoplankton carbohydrate, protein and lipid in the world oceans

Strong spatial variability of the annually-averaged χ_{carbo}, χ_{prot}, χ_{lipid}, carbohydrate, protein and lipid are found over the world's oceanic biomes, for the period of study (Fig. 3). The magnitude of χ_{carbo} varies from <5 in the high-chlorophyll coastal waters and large parts of the northern latitudes beyond 40 degree north (Fig. 3a,c), to >9 in the open oceans and Case I waters (Fig. 3c). Similarly, χ_{prot} (Fig. 3e) or χ_{lipid} (Fig. 3g) vary, respectively, from <15 or <10 in the coastal waters and northern latitudes, to >45 or >30, respectively, in the open oceans and Case I waters. These results generally reflect that the oceanographic regions dominated by large and small phytoplankton are respectively represented by low and high values of χ_{carbo}, χ_{prot} or χ_{lipid}. In the Atlantic and Pacific subtropical gyres, despite the high magnitudes of χ_{carbo}, χ_{prot} and χ_{lipid}, the concentrations of carbohydrate, protein and lipid are typically low (< 0.5, 1.0 and 1.0 $\mathrm{mg\,m^{-3}}$, respectively), and the spatial pattern is similar to the distribution of low chlorophyll. Most of the coastal oceans and Case II waters are generally characterised by higher than 5, 10 and 10 $\mathrm{mg\,m^{-3}}$ of carbohydrate, protein and lipid, respectively, which in places spike beyond 50, 100 and 100 $\mathrm{mg\,m^{-3}}$, respectively (Fig. 3d,f,h). It is noteworthy that some of these very high values may be attributed to the uncertain or erroneous retrievals of chlorophyll and other optical properties in the optically complex water (as also discussed in Section 2.5).

Applying the macromolecular concentration-to-energy conversion factors, i.e, 4.2 $\mathrm{kcal\,g^{-1}}$ for carbohydrate, 4.19 $\mathrm{kcal\,g^{-1}}$ for protein, 9.5 $\mathrm{kcal\,g^{-1}}$ for lipid (Finkel et al., 2016*a*; Hitchcock, 1982), the chemical-energy values of the surface-ocean phytoplankton can be computed (Fig. 3b,d,f,h). The annual average of the phytoplankton energy-value is generally less than 0.1 kJ per $\mathrm{m^{-3}}$ of ocean water in the subtropical gyres, but goes up to 0.5–1.0 kJ per $\mathrm{m^{-3}}$ in parts of the equatorial, northern and southern latitudes, and beyond 10 kJ per $\mathrm{m^{-3}}$ in certain coastal and optically complex waters (Fig. 3b).

3.4 Size-partitioned phytoplankton carbohydrate, protein and lipid in the world oceans

Picoplankton contributions to carbohydrate (in the range [0.1, 1.0] $\mathrm{mg\,m^{-3}}$), protein (in the range [1.0 5.0] $\mathrm{mg\,m^{-3}}$) or lipid (in the range [0.5, 3.0] $\mathrm{mg\,m^{-3}}$) dominate over the contributions of nanoplankton and microplankton in the open oceans and equatorial gyres (Fig. 4a,d,g).

In the northern latitudes beyond 40 degrees and in coastal waters, microplankton contributions to carbohydrate, protein and lipid are higher than those of picoplankton and nanoplankton, with approximate ranges [2.5, 10], [2.0 25] and [0.5, 5.0] mg m^{-3}, respectively (Fig. 4c,f,i). Nanoplankton contributions are generally in the range [1, 3] mg m^{-3} of carbohydrate, [1, 5] mg m^{-3} of protein and [1, 5] mg m^{-3} of lipid, respectively (Fig. 4b,e,h), except in the oligotrophic gyres, where all the concentrations reduce to less than 0.05 mg m^{-3} (Fig. 4b,e,h).

3.5 Macromolecular concentrations in Longhurst provinces

The geographical variations of carbohydrate, protein and lipid in the world oceans can be inferred from their regionally-binned concentrations in the Longhurst biogeographical provinces (Longhurst, 1995, 1998). Given that the ocean-colour data from satellites are inadequate (and may be more erroneous) in the polar regions over most of the year, the estimates from the polar provinces (6 out of 54 Longhurst provinces) are excluded from further discussion. For the remaining 48 provinces, the spatial estimates of χ_{carbo}, χ_{prot}, χ_{lipid} and the concentrations carbohydrate, protein and lipid are computed from their corresponding annually-averaged global maps (Fig. 5). These provinces include 14 Westerlies (NADR, GFST, NASW, MEDI, NASE, PSAE, PSAW, KURO, NPPF, NPSW, TASM, SPSG, SSTC, SANT), 12 Trades (NATR, WTRA, ETRA, SATL, CARB, MONS, ISSG, NPTG, PNEC, PEQD, WARM, ARCH) and 22 Coastal (NECS, CNRY, GUIN, GUIA, NWCS, BRAZ, FKLD, BENG, EAFR, REDS, ARAB, INDE, INDW, AUSW, ALSK, CCAL, CAMR, CHIL, CHIN, SUND, AUSE, NEWZ) provinces (full names of the provinces are given in Supplementary Table S2, and the descriptions in Longhurst, 1995, 1998). The Westerlies, Trades and Coastal provinces are shown in Fig. 2a.

Spatial variability of the estimates in the Coastal provinces are found to be higher than those in the Westerlies or Trades provinces (Fig. 5), with the lowest variability in the Westerlies provinces (Fig. 3-5), reflecting that coastal upwellings would strongly influence the distribution of phytoplankton macromolecules (similar to chlorophyll distribution). The spatial medians of χ_{carbo}, χ_{prot} and χ_{lipid} are lowest (5.69, 13.86 and 8.0, respectively) for the NWCS (North-West Atlantic Coastal Shelves) province, and highest (8.95, 45.13, 29.56, respectively) for the NPSW (North Pacific Subtropical Gyre West) province (Fig. 5a,b,c, and Supplementary Table S2). The NECS (North-East Atlantic Coastal Shelves) province is characterised by the highest surface concentrations (Fig. 5d,e,f) of the annually-averaged spatial medians of carbohydrate (9.53 mg m^{-3}), protein (25.2 mg m^{-3}), and lipid (14.81 mg m^{-3}). The lowest surface concentrations (spatial median) of carbohydrate (0.60 mg m^{-3}) and lipid (1.75 mg m^{-3}) are obtained in the NATR (North Atlantic Tropical Gyral) province (Fig. 5d,f, Table S2), whereas, the lowest concentrations of protein (2.11 mg m^{-3}) is obtained in the NPSW (North Pacific Subtropical Gyre West) province (Fig. 5e Table S2), both of which are generally populated by small picoplankton throughout the year.

The size-partitioned estimates also vary considerably across the 48 Longhurst provinces (Table S3, also Fig. 4). The spatial medians of picoplankton carbohydrate, protein and lipid

are lowest (0.13, 1.11 and 0.83 mg m^{-3}, respectively) in the MEDI (Mediterranean Sea, Black Sea) province, and highest (1.87, 13.33 and 9.43 mg m^{-3}, respectively) in the NECS (NE Atlantic Coastal Shelves) province (Table S2, Fig. 4a,d,g). For nanoplankton, the median concentrations vary from their lowest values (0.09, 0.13 and 0.06 mg m^{-3}, respectively) in the WARM (W. Pacific Warm Pool Trades) province, to their highest values (3.34, 7.36 and 3.66 mg m^{-3}, respectively) in the CHIN (China Sea Coastal) province (Table S2, Fig. 4b,e,h). For microplankton, the median concentrations of carbohydrate and protein vary from their lowest values (0.01 mg m^{-3} for both) in the WARM province, to their highest values (3.39 and 3.37 mg m^{-3}, respectively) in the CHIN province; but that for lipid is found to be highest (1.42 mg m^{-3}) in the NECS (NE Atlantic Coastal Shelves) province, and lowest (0.01 mg m^{-3}) in the WARM province (Table S2, Fig. 4c,f,i). Unsurprisingly, the province-wise distribution of the three macomolecular concentrations show spatial patterns generally consistent with our understanding of the biogeography of phytoplankton size structure.

3.6 Global-ocean stocks of phytoplankton macromolecules

The annually-averaged global stocks are: 0.044 Gt of carbohydrate with monthly range [0.041, 0.05] Gt; 0.17 Gt of protein with monthly range [0.155, 0.18] Gt; and 0.108 Gt of lipid with monthly range [0.098, 0.121] Gt (Fig. 6, and Supplementary Table S4). The largest global stocks are obtained in the month of September, which generally matches with the time of phytoplankton bloom in large parts of the equatorial-southern hemisphere (Kostadinov et al., 2017). The smallest stocks are obtained in the month of June, generally after the termination of the spring blooms.

The percentages of the size-partitioned carbohydrate, protein and lipid stocks also vary over the months of the years (Fig. 6). The stocks constitute the lowest percentage of picoplankton carbohydrate ∼46% (equivalent to 0.02 Gt, with monthly range of 43-53%), compared with the percentages of picoplankton protein ∼78% (equivalent to 0.133 Gt, with monthly range of 76-83%), and picoplankton lipid ∼85% (equivalent to 0.092 Gt, with monthly range of 83-88%) (Supplementary Table S4). The stocks further constitute ∼33% of nanoplankton carbohydrate (equivalent to 0.015 Gt, with monthly range of 32-36%), which is considerably higher that the percentages of nanoplankton protein ∼17% (equivalent to 0.028 Gt, with monthly range of 14-18%), and nanoplankton lipid ∼12% (equivalent to 0.013 Gt, with monthly range of 10-13%). Similarly, the percentage of microplankton carbohydrate ∼21% (equivalent to 0.009 Gt, with monthly range of 16-24%) is significantly higher than the percentages of microplankton protein ∼5% (equivalent to 0.009 Gt, with monthly range of 3-7%) and microplankton lipid ∼3% (equivalent to 0.003 Gt, with monthly range of 2-4%). But clearly, for any given macromolecular stock, the largest contribution comes from picoplankton and the smallest from microplankton (Fig. 6).

No previous estimates were available to compare with the stocks of carbohydrate, protein and lipid reported here. However, the carbon-based macromolecular stocks could be viewed in conjunction with the stocks of total phytoplankton biomass (in carbon units), which were

estimated previously from satellite remote sensing (e.g., Behrenfeld et al., 2005; Kostadinov et al., 2016; Roy et al., 2017). For example, the anually-averaged stocks of the total phytoplankton biomass varied between 0.2 GtC to 1.0 GtC depending on the estimation method (e.g., Behrenfeld et al., 2005; Falkowski et al., 1998; Kostadinov et al., 2016; Roy et al., 2017; Stramski et al., 2008; Taylor et al., 2012). The annually-averaged stocks of carbohydrate protein and lipid and their sum total, estimated above, are within this range. Recent studies (Finkel et al., 2016*a*,*b*) also suggested that under 'nutrient-sufficient, exponential growth conditions' the median composition of the dry weight of microalgae contains 15% carbohydrate 32.2% protein and 17.3% lipid. With respect to the most recent satellite-based estimates of phytoplankton biomass (i.e., ∼0.3 GtC, based on Kostadinov et al., 2016; Roy et al., 2017), the percentages of the annually-averaged global stocks (which included both nutrient sufficient and oligotrophic waters) of carbohydrate, protein and lipid are $\sim 15\%$, $\sim 57\%$, $\sim 36\%$, respectively. These preliminary results thus suggests that on a global scale, the relative proportions of carbohydrate in phytoplankton might be more robust than the proportions of protein and lipid. However, direct *in situ* measurements would be required to further validate these results.

3.7 Algorithm uncertainties on global map

The uncertainty propagation maps based on the sensitivity analysis suggest that the relative uncertainties in lipid estimates would be higher than those in protein or carbohydrate for most of the world's productive regions (Fig. 7); but in the less productive oligotrophic waters, the relative uncertainties in all the estimates would be generally comparable. The relative uncertainties in carbohydrate estimates would be within $30-45\,\%$ in most of the upwelling and productive regions and coastal waters, but would reduce to $<15\%$ in the subtropical gyres and oligotrophic waters (Fig. 7a). Similar spatial pattern are obtained for the relative uncertainties in protein and lipid estimates, although the magnitudes of the relative uncertainties would be different. For protein and lipid the relative uncertainties would be $<15\%$ and $<25\%$, respectively, inside the gyres, and between $30-40\,\%$ and $35-50\,\%$, respectively, in major parts of the Northern hemisphere; and but would increase up to $60-64\,\%$ and $65-80\,\%$, respectively, in large parts of the southern ocean and around the overlapping regions of the oligotrophic and eutrophic waters (Fig. 7b,c).

4 Concluding remarks

Although a variety of satellite-based ocean-colour algorithms have already been developed to retrieve chlorophyll-a and its contributions in PFTs and PSCs (e.g., review by Mouw et al., 2017), and phytoplankton carbon (Behrenfeld et al., 2005; Kostadinov et al., 2016; Roy et al., 2017; Sathyendranath et al., 2009), no methodology exists so far to estimate from satellites, the concentrations of macromolecules that essentially determine the energy value of phytoplankton. The bio-optical method presented here would be the first one to compute,

from satellite data, the concentrations of phytoplankton carbohydrate, protein and lipid, and the resultant energy value of phytoplankton on a global scale. In this novel approach, the satellite-derived bio-optical fingerprints of the living phytoplankton combined with allometric relationships are used, which builds on the ocean-colour algorithms recently developed for retrieving phytoplankton cell size, the exponent of the phytoplankton size spectra, phytoplankton carbon and PSCs from satellite (Roy et al., 2013, 2011, 2017). Presented are the first estimates of annually-averaged concentrations of carbohydrate, protein, lipid, and ratios of chlorophyll-a to cellular macromolecular concentrations over the global oceans as well as those for the Longhurst biogeochemical provinces, over the period 1997-2013. Although the current estimates are based on the OC-CCI merged satellite products, by design, the methodology would be equally applicable to ocean-colour data from any other satellite sensor.

Recent studies based on either ocean-colour data (Behrenfeld et al., 2005; Kostadinov et al., 2016; Roy et al., 2017; Stramski et al., 2008) or Earth System models (e.g., CMIP5, Taylor et al., 2012), have attempted to improve the estimates of the stocks of phytoplankton carbon, and have narrowed down the estimation range of the annually-averaged stocks. But unclear is how the total carbon stock partitions into the stocks of essential carbon-based macromolecules in phytoplankton. For example, although the proportions of the macromolecules to dry weight of phytoplankton are reported for ideal nutrient-rich conditions (Finkel et al., 2016*a*,*b*), little in known about those proportions in diverse oceanographic regions where growth conditions deviate from ideal. This study independently estimates the annually-averaged stocks of the three essential phytoplankton macromolecules, and finds that the sum total of these estimates are well within the range of the reported stocks of total phytoplankton carbon. The estimates would be potentially useful for understanding the cellular allocation of carbon to carbohydrate, protein and lipid pools in phytoplankton, both spatially and over time, with implications for trophic transfer models, and higher trophic or fisheries models.

The lack of adequate direct measurements on carbohydrate, protein and lipid overlapping the temporal coverage of the ocean-colour data have restricted rigorous validation of the satellite-derived estimates. Therefore, new *in situ* measurements of phytoplankton macromolecules across various oceanic conditions should be a priority, for increasing the reliability and reducing the bias and uncertainties of the satellite-based estimates. Adequate direct measurement would also allow computation of observation-based uncertainties such as RMSE and bias, pixel-by-pixel, and providing those to the users. The sensitivity analyses carried out here, with assumptions on fixed relative uncertainties of $<30\%$ for the input parameters (following the requirement provided by Global Climate Observing System, GCOS, 2011), have identified oceanographic regions where the estimates would be less (or more) sensitive to relative uncertainties in satellite inputs. But, how the relative uncertainties may alter (reduce or increase), due to regional variations of uncertainties in the input parameters, and how those may impact the estimates of the global stocks, would require further investigations. The sensitivity analyses however have shown promise that the estimation errors could reduce, as the retrievals of satellite-based IOPs become more accurate. Finally, due to the constraints of inadequate *in situ* validation data, and large uncertainties and biases in the optically complex waters, arising from the presence of high concentration of coloured-dissolved organic matters,

sediments, clouds and ice, the current estimates may be less reliable in coastal waters and high latitudes, than those in open oceans. So, the applicability and reliability of the estimates to optically complex waters would also be subject to further investigations, possibly including improved satellite inputs, as the satellite era enters into higher temporal and spatial resolution.

Acknowledgements

This work was a part of SR's ongoing research at the University of Reading and was supported by an International Exchanges Award from the Royal Society of London. The satellite data were obtained freely from the Ocean Colour Climate Change Initiative programme - the project team is acknowledged for generating and sharing the merged datasets on chlorophyll and inherent optical properties. The mission scientists and Principal Investigators and everyone associated with compilation of the marine biodiversity database and MICOC data used here are also acknowledged for making these data freely available. Helpful comments and constructive suggestions from the reviewers and editor improved the paper.

References

Behrenfeld, M. J., Boss, E., Siegel, D. A. and Shea, D. M. (2005), 'Carbon-based ocean productivity and phytoplankton physiology from space', *Global biogeochemical cycles* **19**(1), GB1006.

Breteler, W. K., Schogt, N. and Rampen, S. (2005), 'Effect of diatom nutrient limitation on copepod development: role of essential lipids', *Marine Ecology Progress Series* **291**, 125–133.

Brewin, R. J., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E. et al. (2015), 'The ocean colour climate change initiative: Iii. a round-robin comparison on in-water bio-optical algorithms', *Remote Sensing of Environment* **162**, 271–294.

Budge, S. M., Devred, E., Forget, M.-H., Stuart, V., Trzcinski, M. K., Sathyendranath, S. et al. (2014), 'Estimating concentrations of essential omega-3 fatty acids in the ocean: supply and demand', *ICES Journal of Marine Science: Journal du Conseil* **71**(7), 1885–1893.

Chisti, Y. (2007), 'Biodiesel from microalgae', *Biotechnology advances* **25**(3), 294–306.

Falkowski, P. (2012), 'The power of plankton', *Nature* **483**, S17–S20.

Falkowski, P. G., Barber, R. T. and Smetacek, V. (1998), 'Biogeochemical controls and feedbacks on ocean primary production', *Science* **281**(5374), 200–206.

Field, C. B., Behrenfeld, M. J., Randerson, J. T. and Falkowski, P. (1998), 'Primary production of the biosphere: integrating terrestrial and oceanic components', *Science* **281**(5374), 237–240.

Finkel, Z., Follows, M. and Irwin, A. (2016*a*), 'Size-scaling of macromolecules and chemical energy content in the eukaryotic microalgae', *Journal of Plankton Research* **38**(5), 1151–1162.

Finkel, Z., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I. and Irwin, A. J. (2016*b*), 'Phylogenetic diversity in the macromolecular composition of microalgae', *PloS one* **11**(5), e0155977.

GCOS, G. (2011), 'Systematic observation requirements for satellite-based products for climate. 2011 update supplemetnatl details to the satellite 39 based component og the implementation plan for the global observing system for climate in support of the unfccc (2010 update)', Tech. rep., World Meteorological Organisation (WMO), 7 bis, avenue de la Paix, CH- 1211 Geneva 2, Switzerland.

Geider, R. and La Roche, J. (2002), 'Redfield revisited: variability of c: N: P in marine microalgae and its biochemical basis', *European Journal of Phycology* **37**(1), 1–17.

Goldman, J. C. and Caron, D. A. (1985), 'Experimental studies on an omnivorous microflagellate: implications for grazing and nutrient regeneration in the marine microbial food chain', *Deep Sea Research Part A. Oceanographic Research Papers* **8**, 899–915.

Hitchcock, G. L. (1982), 'A comparative study of the size-dependent organic composition of marine diatoms and dinoflagellates', *Journal of plankton research* **4**(2), 363–377.

IOCCG (2000), Remote sensing of ocean colour in coastal, and other optically-complex, waters, Technical Report 3, Dartmouth, NS.

IOCCG (2006), Remote sensing of inherent optical properties: Fundamentals, tests of algorithms, and applications, Technical Report 5, Dartmouth, Canada.

IOCCG, ed. (2014), *Phytoplankton functional types from Space.*, number 15, International Ocean-Colour Coordinating Group, Reports of the International Ocean-Colour Coordinating Group (IOCCG).

Jónasdóttir, S. (1994), 'Effects of food quality on the reproductive success of acartia tonsa and acartia hudsonica: laboratory observations', *Marine Biology* **121**(1), 67–81.

Kostadinov, T., Milutinovic, S., Marinov, I. and Cabré, A. (2016), 'Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution', *Ocean Science Discussions* **12**, 561–575.

584 Kostadinov, T. S., Cabré, A., Vedantham, H., Marinov, I., Bracher, A., Brewin, R. J.
585 et al. (2017), 'Inter-comparison of phytoplankton functional type phenology metrics de-
586 rived from ocean color algorithms and earth system models', *Remote Sensing of Environ-*
587 *ment* **190**, 162–177.

588 Litzow, M. A., Bailey, K. M., Prahl, F. G. and Heintz, R. (2006), 'Climate regime shifts and
589 reorganization of fish communities: the essential fatty acid limitation hypothesis', *Marine*
590 *Ecology Progress Series* **315**, 1–11.

591 Longhurst, A. R. (1995), 'Seasonal cycles of pelagic production and consumption', *Progress*
592 *in oceanography* **36**(2), 77–167.

593 Longhurst, A. R. (1998), *Ecological Geography of the Sea*, Academic Press.

594 Marañón, E. (2008), 'Inter-specific scaling of phytoplankton production and cell size in the
595 field', *Journal of Plankton Research* **30**(2), 157–163.

596 Marañón, E., Cermeño, P., López-Sandoval, D. C., Rodríguez-Ramos, T., Sobrino, C., Huete-
597 Ortega, M. et al. (2013), 'Unimodal size scaling of phytoplankton growth and the size
598 dependence of nutrient uptake and use', *Ecology letters* **16**(3), 371–379.

599 Marañón, E., Cermeno, P., Rodrıguez, J., Zubkov, M. V. and Harris, R. P. (2007), 'Scaling of
600 phytoplankton photosynthesis and cell size in the ocean', *Limnol. Oceanogr* **52**(5), 2190–
601 2198.

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

605 Martiny, A. C., Pham, C. T., Primeau, F. W., Vrugt, J. A., Moore, J. K., Levin, S. A. et al.
606 (2013), 'Strong latitudinal patterns in the elemental ratios of marine plankton and organic
607 matter', *Nature Geoscience* **6**(4), 279–283.

608 McCave, I. (1984), 'Size spectra and aggregation of suspended particles in the deep ocean',
609 *Deep Sea Research Part A. Oceanographic Research Papers* **31**(4), 329–352.

610 Menden-Deuer, S. and Lessard, E. J. (2000), 'Carbon to volume relationships for dinoflagel-
611 lates, diatoms, and other protist plankton', *Limnology and Oceanography* **45**(3), 569–579.

612 Moal, J., Martin-Jezequel, V., Harris, R., Samain, J.-F. and Poulet, S. (1987), 'Interspe-
613 cific and intraspecific variability of the chemical-composition of marine-phytoplankton',
614 *Oceanologica acta* **10**(3), 339–346.

615 Mouw, Colleen, B., Hardman-Mountford, N., Alvain, S., Bracher, A., Brewin, R. J. W.,
616 Bricaud, A. et al. (2017), 'A consumer's guide to satellite remote sensing of multiple phy-
617 toplankton groups in the global ocean', *Frontiers in Marine Science* **4**(41).

618 Müller, D., Krasemann, H., Brewin, R. J., Brockmann, C., Deschamps, P.-Y., Doerffer, R.
619 et al. (2015), 'The ocean colour climate change initiative: Ii. spatial and temporal ho-
620 mogeneity of satellite data retrieval due to systematic effects in atmospheric correction
621 processors', *Remote Sensing of Environment* **162**, 257–270.

622 Opute, F. I. (1974), 'Lipid and fatty-acid composition of diatoms', *Journal of Experimental*
623 *Botany* pp. 823–835.

624 Peters, R. H. (1983), *The Ecological Implications of Body Size*, Cambridge University Press,
625 Cambridge.

626 Pethybridge, H. R., Parrish, C. C., Morrongiello, J., Young, J. W., Farley, J. H., Gunasekera,
627 R. M. et al. (2015), 'Spatial patterns and temperature predictions of tuna fatty acids:
628 tracing essential nutrients and changes in primary producers', *PloS one* **10**(7), e0131598.

629 Reynolds, R., Stramski, D., Wright, V. and Woźniak, S. (2010), 'Measurements and charac-
630 terization of particle size distributions in coastal waters', *Journal of Geophysical Research:*
631 *Oceans* **115**(C8), C08024.

632 Roy, S., Alam, S. and Chattopadhyay, J. (2005), 'Role of nutrient bound of prey on the
633 dynamics of predator-mediated competitive-coexistence', *BioSystems* **82**(2), 143–153.

634 Roy, S. and Chattopadhyay, J. (2007*a*), 'Enrichment and ecosystem stability: effect of toxic
635 food', *BioSystems* **90**(1), 151–160.

636 Roy, S. and Chattopadhyay, J. (2007*b*), 'Enrichment and stability: A phenomenological cou-
637 pling of energy value and carrying capacity', *BioSystems* **90**(2), 371–378.

638 Roy, S., Sathyendranath, S., Bouman, H. and Platt, T. (2013), 'The global distribution of
639 phytoplankton size spectrum and size classes from their light-absorption spectra derived
640 from satellite data', *Remote Sensing of Environment* **139**, 185–197.

641 Roy, S., Sathyendranath, S. and Platt, T. (2011), 'Retrieval of phytoplankton size from bio-
642 optical measurements: theory and applications', *Journal of The Royal Society Interface*
643 **8**(58), 650–660.

644 Roy, S., Sathyendranath, S. and Platt, T. (2017), 'Size-partitioned phytoplankton carbon
645 and carbon-to-chlorophyll ratio from ocean-colour by an absorption-based bio-optical algo-
646 rithm', *Remote Sensing of Environment* **194**, 177–189.

647 Sal, S., López-Urrutia, Á., Irigoien, X., Harbour, D. S. and Harris, R. P. (2013), 'Marine
648 microplankton diversity database', *Ecology* **94**(7), 1658–1658.

649 Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H. et al. (2009),
650 'Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea', *Marine Ecology*
651 *Progress Series* **383**(7), 73–84.

Schmidtko, S., Johnson, G. C. and Lyman, J. M. (2013), 'MIMOC: A global monthly isopycnal upper-ocean climatology with mixed layers', *Journal of Geophysical Research: Oceans* **118**(4), 1658–1672.

Sheldon, R., Prakash, A. and Sutcliffe, W. (1972), 'The size distribution of particles in the ocean', *Limnology and oceanography* **17**(3), 327–340.

Shin, K., Jang, M.-C., Jang, P.-K., Ju, S.-J., Lee, T.-K. and Chang, M. (2003), 'Influence of food quality on egg production and viability of the marine planktonic copepod acartia omorii', *Progress in Oceanography* **57**(3), 265–277.

Sieburth, J. M., Smetacek, V. and Lenz, J. (1978), 'Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1', *Limnology and Oceanography* **23**(6), 1256–1263.

Sterner, R. W. and Elser, J. J. (2002), *Ecological stoichiometry: the biology of elements from molecules to the biosphere*, Princeton University Press.

Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R. et al. (2008), 'Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern south pacific and eastern atlantic oceans', *Biogeosciences* **5**(1), 171–201.

Strathmann, R. (1967), 'Estimating the organic carbon content of phytoplankton from cell volume or plasma volume', *Limnology and Oceanography* **12**(3), 411–418.

Taylor, K. E., Stouffer, R. J. and Meehl, G. A. (2012), 'An overview of cmip5 and the experiment design', *Bulletin of the American Meteorological Society* **93**(4), 485–498.

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A. and Marty, J.-C. (2001), 'Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern mediterranean sea during winter', *Journal of Geophysical Research: Oceans (1978–2012)* **106**(C9), 19939–19956.

Werdell, P. J., Bailey, S. W., Franz, B. A., Harding, L. W., Feldman, G. C. and McClain, C. R. (2009), 'Regional and seasonal variability of chlorophyll-a in chesapeake bay as observed by seawifs and modis-aqua', *Remote Sensing of Environment* **113**(6), 1319–1330.1

Figure captions

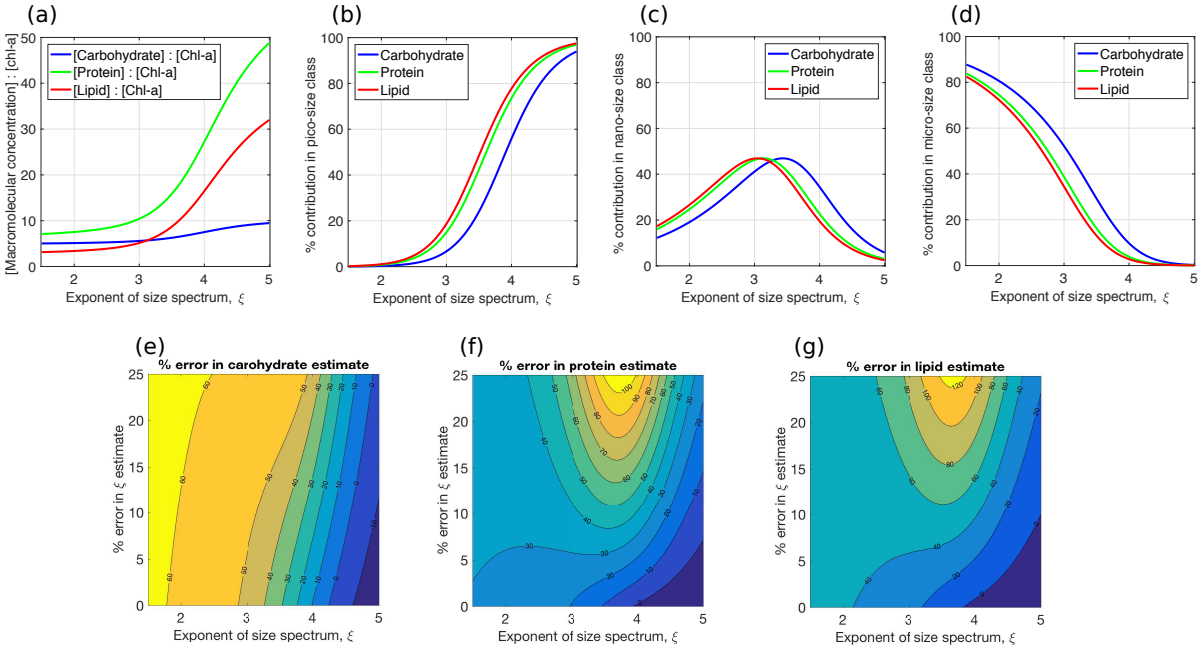

Figure 1: (a) Carbohydrate-to-chlorophyll (χ_{carbo}), protein-to-chlorophyll (χ_{prot}) and lipid-to-chlorophyll (χ_{lipid}) ratios of the mixed-phytoplankton population derived (using Eq. 3) as functions of the exponent of the phytoplankton size spectrum (ξ). (b)-(d) Size-partitioned carbohydrate, protein and lipid proportions in: (b) picoplankton, (c) nanoplankton and (d) microplankton, derived using Eq. (8). (e)-(g) Algorithm-based relative uncertainties in the estimates of: (e) carbohydrate, (f) protein and (g) lipid, quantified as a joint function of the relative uncertainties in ξ, a_M and b_M (see, Section 2.5). The 95% confidence levels for the allometric parameters reported in Finkel et al. (2016*a*) are considered for computing the % uncertainties in the parameters with respect to their reported means, along with a range of 0-25% relative uncertainty in ξ (following Roy et al. (2013)).

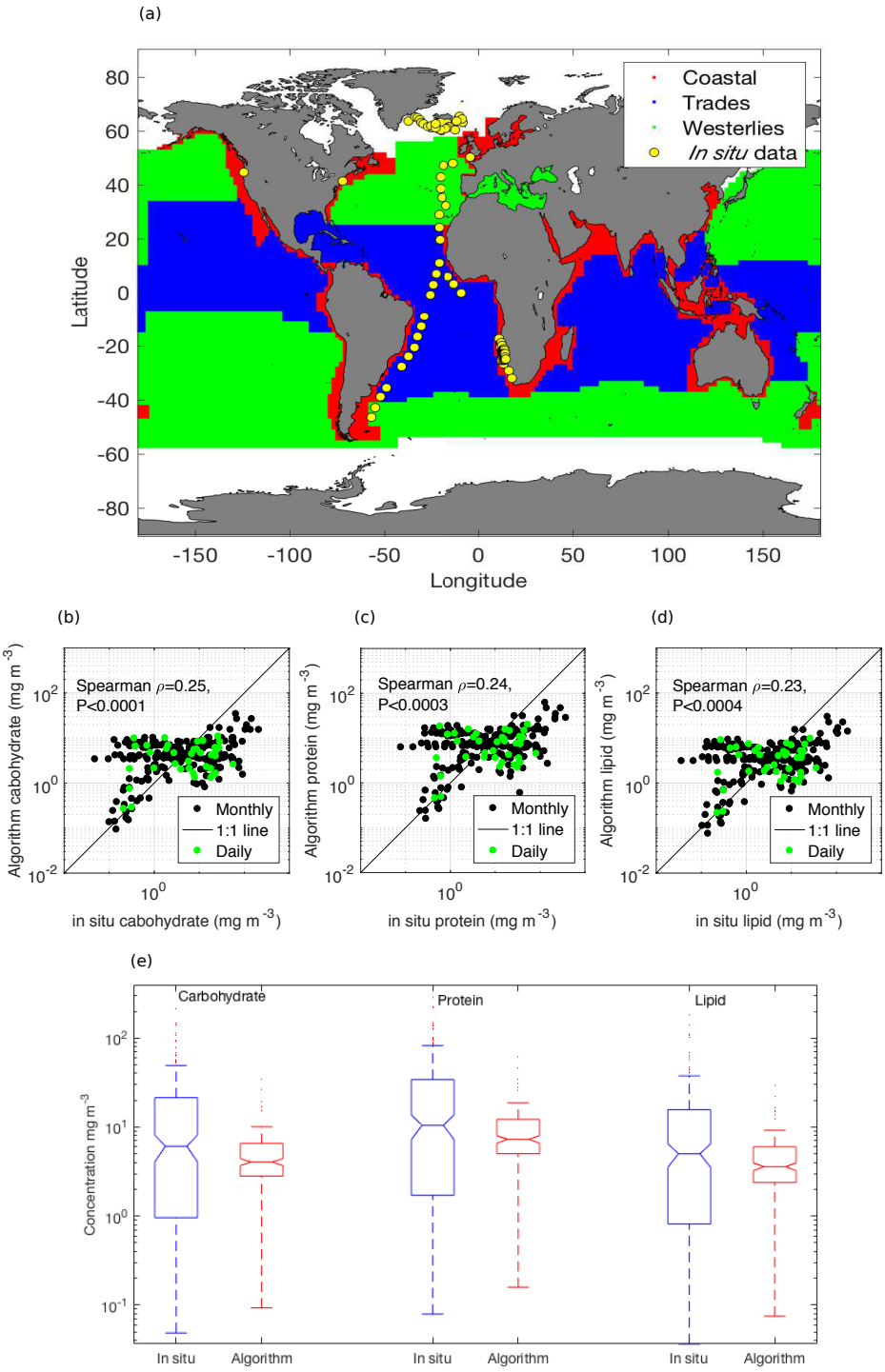

Figure 2: (a) Geographic locations of the *in situ* samples (yellow dots) used from the marine biodiversity database (Sal et al., 2013); this subset overlapped with the temporal coverage of satellite data, and were considered for computing phytoplankton carbohydrate, protein and lipid using species size and cell abundances, and by applying the allometric relationships reported in Finkel et al. (2016*a*). The Westerlies, Trades and Coastal Longhurst provinces are shown in different colours. (b)-(d) Satellite match-ups from daily (green dots) and monthly (black dots) images were considered for comparing the satellite-derived (b) carbohydrate, (c) protein and (d) lipid with the *in situ* estimates. (e) Box-plots comparing the estimates from *in situ* with satellite based on the current method.

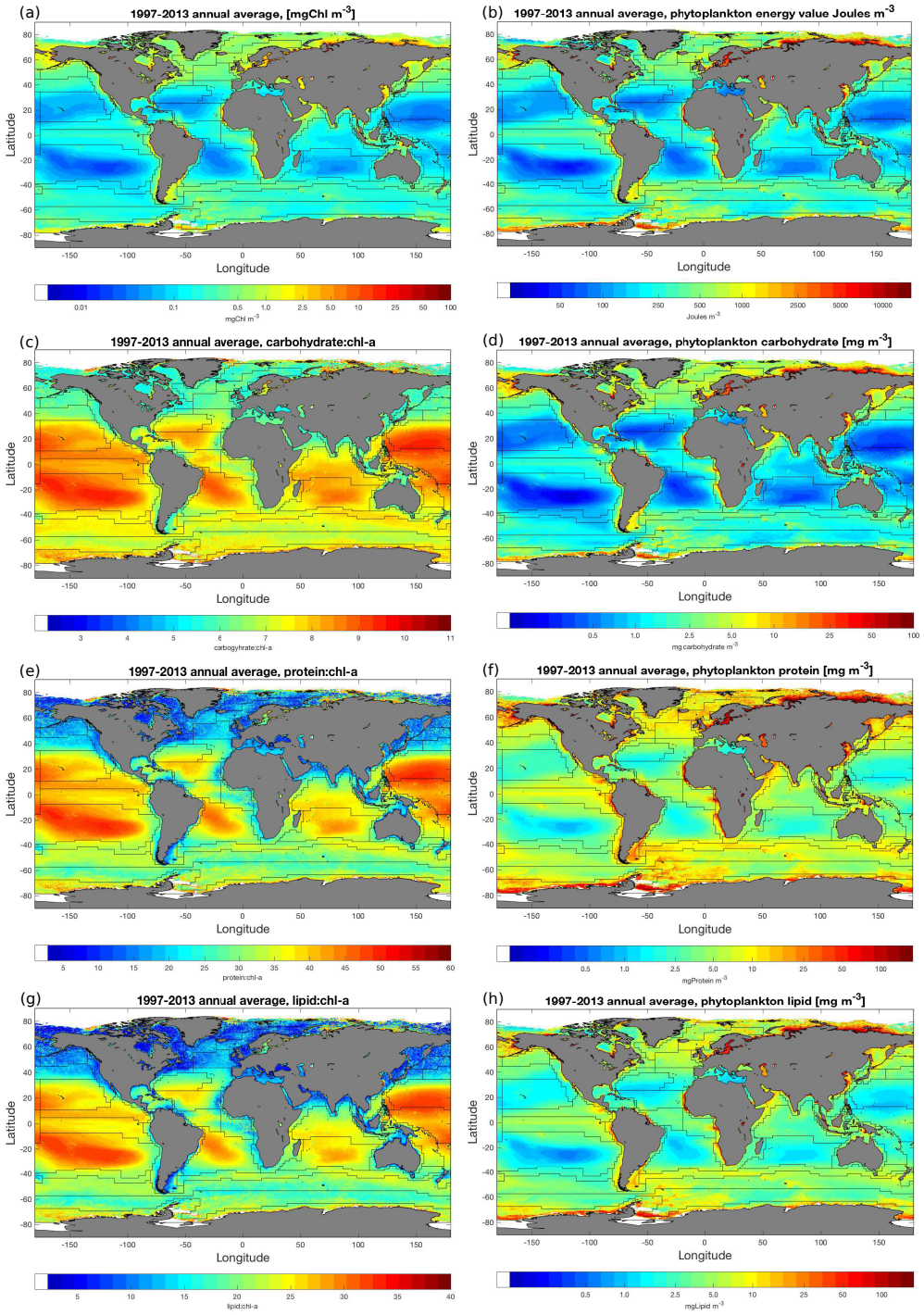

Figure 3: Distributions of the annually-averaged surface concentrations of macromolecules and energy value of phytoplankton over 1997-2013. Overlaid on the global maps are thin black lines representing the boundaries of the Longhurst biogeographical provinces (Longhurst, 1995, 1998). Annual averages of (a) surface chlorophyll in [mg m^{-3}]; (b) chemical energy value of phytoplankton in [Joules m^{-3}] as a combinations of the estimated carbohydrate, protein, lipid; (c) carbohydrate to chlorophyll ratio (dimensionless); (d) concentration of carbohydrate in [mg m^{-3}]; (e) protein to chlorophyll ratio (dimensionless); (f) concentration of protein in [mg m^{-3}]; (g) lipid to chlorophyll ratio (dimensionless); (h) concentration of lipid in [mg m^{-3}], computed based on the methodology described in Section 2.

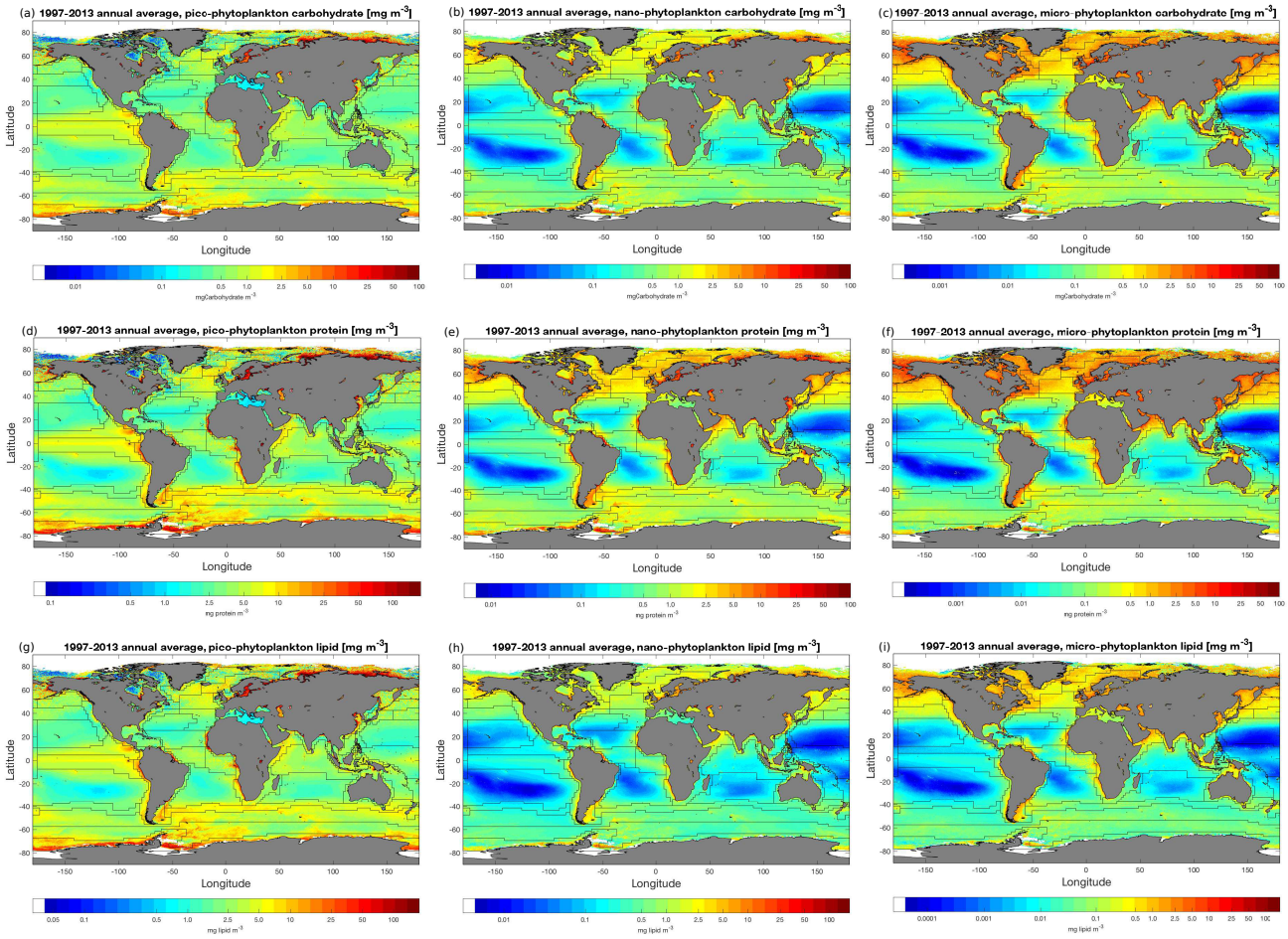

Figure 4: Annually-averaged surface macromolecular concentrations [mg m^{-3}] in picoplankton, nanoplankton, and microplankton over 1997-2013: (a) picoplankton carbohydrate, (b) nanoplankton carbohydrate, (c) microplankton carbohydrate; (d) picoplankton protein, (e) nanoplankton protein, (f) microplankton protein; and (g) picoplankton lipid, (h) nanoplankton lipid, (i) microplankton lipid, computed based on the methodology described in Section 2. Overlaid on the global maps are thin black lines representing the boundaries of the Longhurst biogeographical provinces (Longhurst, 1995, 1998).

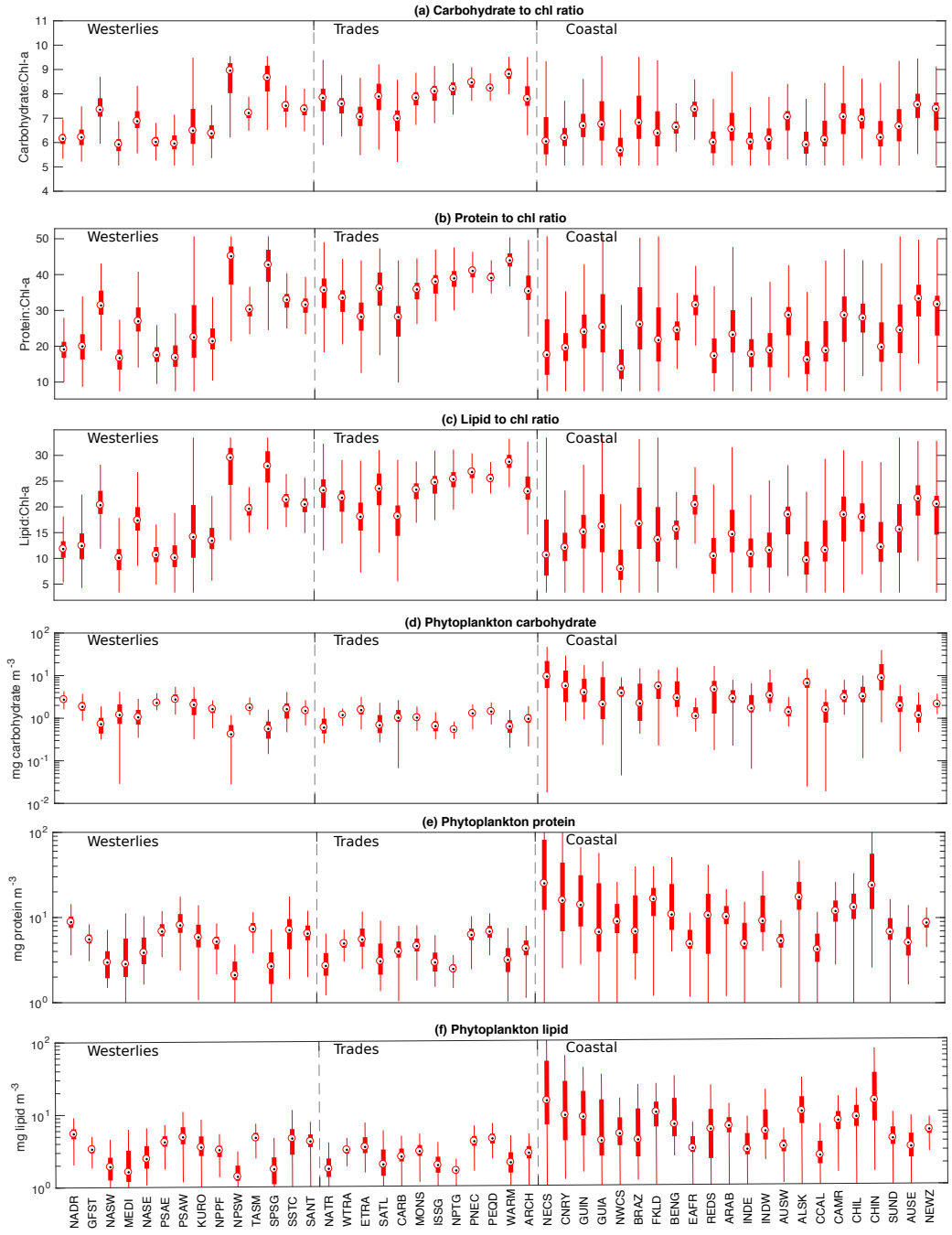

Figure 5: Annually-averaged surface macromolecular composition within Longhurst biogeographical provinces (Longhurst, 1995) computed over 1997-2013. Box plots with annual median (black dots), interquartile ranges (thick red bar), and ranges (thin whiskers) for (a) carbohydrate-to-chlorophyll ratio (χ_{carbo}), (b) protein-to-chlorophyll ratio (χ_{prot}), (c) lipid-to-chlorophyll ratio (χ_{lipid}), (d) carbohydrate ($\mathrm{mg\,m^{-3}}$), (e) protein ($\mathrm{mg\,m^{-3}}$), and (f) lipid ($\mathrm{mg\,m^{-3}}$), are shown for 48 Longhurst provinces. The provinces include 14 Westerlies (NADR, GFST, NASW, MEDI, NASE, PSAE , PSAW, KURO, NPPF, NPSW, TASM, SPSG, SSTC, SANT), 12 Trades (NATR, WTRA, ETRA, SATL, CARB, MONS, ISSG, NPTG, PNEC, PEQD, WARM, ARCH) and 22 Coastal (NECS, CNRY, GUIN, GUIA, NWCS, BRAZ, FKLD, BENG, EAFR, REDS, ARAB, INDE, INDW, AUSW, ALSK, CCAL, CAMR, CHIL, CHIN, SUND, AUSE, NEWZ) provinces. The provinces within Westerlies, Trades and Coastal are arranged from north to south as they appear in the Longhurst's original list. Descriptions of the provinces can be found in Longhurst (1995, 1998), and the full names of the provinces along with the plotted median values of the annual averages are given in Table S1.

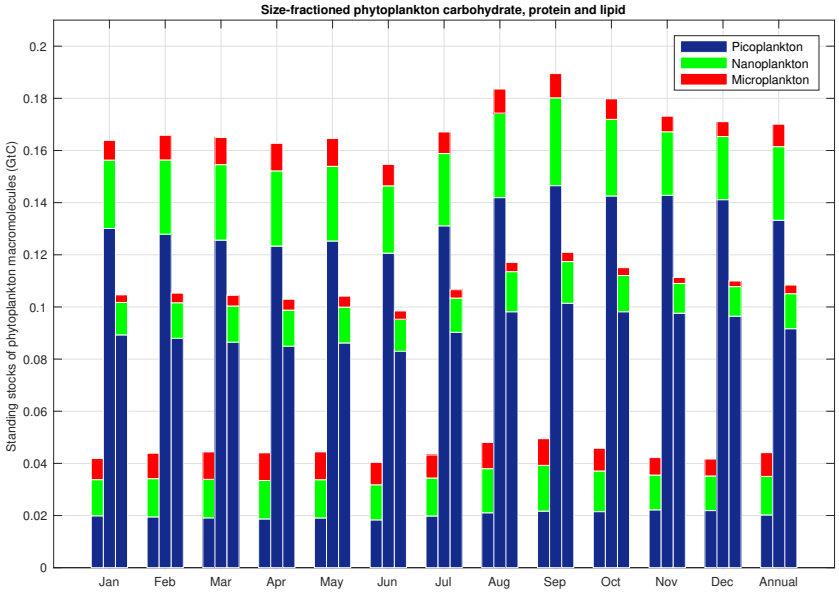

Figure 6: Annually-averaged macromolecular compositions for three phytoplankton size classes over 1997-2013. Grouped bars represent the monthly and annual stocks of the total (height of each bar) and size-partitioned (blue - picoplankton fraction, green - nanoplankton fraction, and red - microplankton fraction) estimates of carbohydrate (first bar in each group), protein (second bar in each group) and lipid (third bar in each group), computed from the surface concentrations through integrations over the mixed-layer depths. All concentrations are expressed in gigatonnes (Gt).

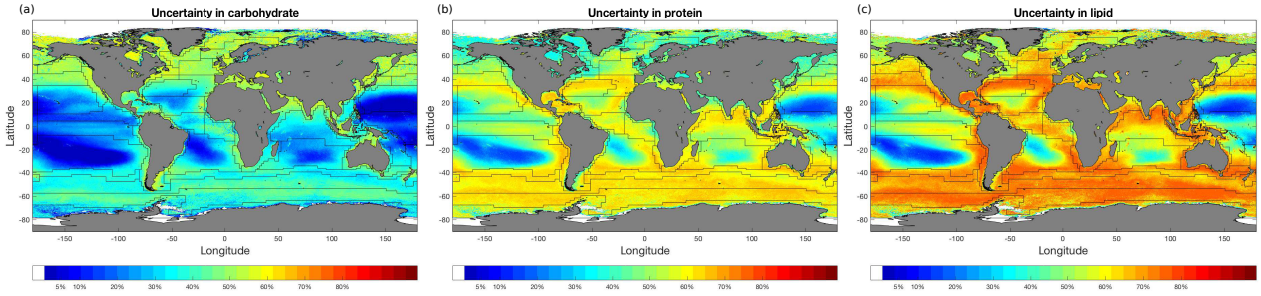

Figure 7: Algorithm uncertainty maps corresponding to the estimates of (a) phytoplankton carbohydrate, (b) phytoplankton protein and (c) phytoplankton lipid based on the sensitivity analysis in Section 2.5. Annually-averaged uncertainties in estimating the surface concentrations of carbohydrate, protein and lipid are shown for an overall relative uncertainty of 25% in ξ retrievals combined with 95% confidence intervals of the allometric parameters reported by Finkel et al. (2016*a*).

Table 1: Summary of overall uncertainties (mean with ranges) in carbohydrate, protein and lipid estimates as a function of uncertainties in ξ and allometric parameters, as shown in Fig. (1e,f,g)

	$\xi < 3.25$	$3.25 \leq \xi \leq 4.5$	$\xi > 4.5$	
$\Delta\xi/\xi \leq 15\%$	56% (41–62%)	25% (0–52%)	0.3% (0–7%)	Carbohydrate
	31% (15–55%)	29% (0–66%)	7% (0–30%)	Protein
	44% (19–70%)	32% (0–77%)	5% (0–30%)	Lipid
$\Delta\xi/\xi > 15\%$	59% (52–62%)	40% (8–60%)	4% (0–18%)	Carbohydrate
	42% (32–83%)	74% (31–108%)	29% (10–55%)	Protein
	58% (46–105%)	85% (31–126%)	27% (7–59%)	Lipid