

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

Article

Accepted Version

Roy, S. (2018) Distributions of phytoplankton carbohydrate, protein and lipid in the world oceans from satellite ocean colour. *The ISME Journal*, 12 (6). pp. 1457-1472. ISSN 1751-7370 doi: <https://doi.org/10.1038/s41396-018-0054-8> Available at <http://centaur.reading.ac.uk/74681/>

It is advisable to refer to the publisher's version if you intend to cite from the work.

To link to this article DOI: <http://dx.doi.org/10.1038/s41396-018-0054-8>

Publisher: Nature Publishing Group

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

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

4 Shovonlal Roy\*<sup>1</sup>

5 <sup>1</sup>Department of Geography and Environmental Science & School of  
6 Agriculture, Policy and Development, University of Reading, Whiteknights,  
7 Reading RG6 6AB, U.K

8 January 26, 2018

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

10

11

12 **Running title:** Satellite-derived carbohydrate, protein and lipid

---

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

13

**Abstract**

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

Energy value of phytoplankton regulates the growth of higher trophic species, affecting the trophic 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.

## 43 1 Introduction

44 The autotrophic phytoplankton species in the upper ocean, constituting less than 1% of the  
45 entire photosynthetic biomass on the globe, are responsible not only for  $\sim 50\%$  of the global an-  
46 nual carbon-fixation (Falkowski, 2012; Field et al., 1998), but also for providing life-support to  
47 marine food-webs through its trophic connections. In addition to their biomass and species  
48 composition, the cellular macromolecular contents and energy value of phytoplankton can  
49 strongly impact the trophic balance within a marine ecosystem, e.g., by directly impacting  
50 the developmental stages of grazers, and influencing the trophic-energy flow affecting the  
51 production of higher trophic species (Breteler et al., 2005; Jónasdóttir, 1994; Litzow et al.,  
52 2006; Shin et al., 2003). The stoichiometric ratio, i.e., the relative elemental composition of  
53 carbon, nitrogen and phosphorous in phytoplankton is known to vary with phytoplankton  
54 assemblages across resource gradients in marine biomes (Geider and La Roche, 2002; Martiny  
55 et al., 2013). Stoichiometric variations alter the nutritional quality of phytoplankton as food  
56 to the grazers (Goldman and Caron, 1985; Sterner and Elser, 2002); and variations in nutri-  
57 ent bound or energy value of phytoplankton affect the stability and oscillatory dynamics of  
58 producer-grazer interactions (e.g., Roy et al., 2005; Roy and Chattopadhyay, 2007*a,b*). It is,  
59 therefore, imperative to monitor the variations in cellular macromolecular contents of marine  
60 phytoplankton, on local, regional and global scales. In this context, possibilities of having  
61 satellite-based estimates would be invaluable, given that *in situ* observations are often infre-  
62 quent, and inadequate for monitoring over large spatial scales. Moreover, conducting *in situ*  
63 measurements of the macromolecular contents of phytoplankton in the global ocean, would  
64 be extremely time consuming and considerably expensive.

65 Over the last two decades, several satellite-based methods have been developed to extend  
66 our capabilities from routinely estimating chlorophyll concentration, to distinguishing phyto-  
67 plankton functional types (PFTs), in terms of the proportions of chlorophyll either in major  
68 taxonomic groups, or in phytoplankton size classes (PSCs) (for more details, see, IOCCG,  
69 2014; Mouw et al., 2017). Some progress has also been made to estimate phytoplankton car-  
70 bon (Behrenfeld et al., 2005; Kostadinov et al., 2016; Roy et al., 2017; Sathyendranath et al.,  
71 2009), and carbon-based classification of PSCs (Kostadinov et al., 2016; Roy et al., 2017),  
72 from ocean colour. However, strong variations in phytoplankton cellular carbon and carbon-  
73 based macromolecules, with taxa, cell morphology, and environmental conditions such as  
74 ambient light and available nutrient (Hitchcock, 1982; Marañón, 2008; Marañón et al., 2013;  
75 Menden-Deuer and Lessard, 2000; Strathmann, 1967), impose additional layers of difficul-  
76 ties in converting the satellite-derived estimates of chlorophyll or carbon-to-macromolecular  
77 concentrations. Certain phytoplankton macromolecules, such as cellular fatty acids, strongly  
78 vary (e.g., between 1% and 85%, Chisti, 2007), not only among algal groups and species, but  
79 also within a specific algal group, e.g., diatoms under different culture conditions (Opute,  
80 1974). In addition to laboratory cultures, essential fatty acids in phytoplankton have also  
81 been reported to vary with oceanographic conditions, such as sea-surface temperature and  
82 chlorophyll-a, on regional scales (e.g., Budge et al., 2014; Pethybridge et al., 2015). However,  
83 progress is yet to be made to estimate the variations in total phytoplankton lipid from satellite

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

90 In this paper, the cellular macromolecular contents of marine phytoplankton, in partic-  
91 ular, the concentrations of carbohydrate, protein and lipid are estimated on a global scale,  
92 for the first time, based on ocean-colour data from satellite remote-sensing. To do so, a  
93 novel method is derived that utilises light-absorption coefficients of phytoplankton ( $a_{ph}$ ) - an  
94 inherent optical property (IOP) retrievable from ocean colour (e.g., IOCCG, 2006), coupled  
95 with allometric relationships between phytoplankton cells and their cellular macromolecular  
96 contents, reported in the literature (Hitchcock, 1982; Marañón, 2008; Marañón et al., 2013,  
97 2007; Menden-Deuer and Lessard, 2000; Moal et al., 1987; Peters, 1983; Strathmann, 1967).  
98 The method builds on a semi-analytical algorithm for retrieving the exponent of the phyto-  
99 plankton size spectrum from satellite ocean colour, developed recently by Roy et al. (2013,  
100 2011). The concentrations of the total macromolecular contents are further partitioned ac-  
101 cording to their contributions in three bulk PSCs, namely, picoplankton, nanoplankton and  
102 microplankton. The estimates are obtained over the global ocean, and for different marine  
103 biomes represented by Longhurst oceanographic provinces (Longhurst, 1995, 1998). Further,  
104 insights on the estimation uncertainties are provided through detailed sensitivity analyses,  
105 highlighting the possibilities of further improvements of the estimates, with the expectation  
106 that the input satellite data would further improve, as the satellite era enters into higher  
107 temporal and spatial resolution.

## 108 2 Methodology

### 109 2.1 Satellite validation

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

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

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

129 Whilst direct measurements on carbohydrate, protein and lipid were unavailable, large  
130 datasets on *in situ* phytoplankton abundance were available, e.g., those compiled in ma-  
131 rine biodiversity database (Sal et al., 2013), which included phytoplankton cell counts from  
132 samples collected in different oceanographic cruises between 1992 and 2002, partly covering  
133 the satellite period. Owing to the constraints on direct measurements, a validation exer-  
134 cise was attempted by converting the *in situ* data on phytoplankton abundance (Sal et al.,  
135 2013) into estimates of phytoplankton macromolecular concentrations, using allometric re-  
136 lationships from the literature (Finkel et al., 2016a). To do so, a subset of phytoplankton  
137 abundance data (Sal et al., 2013) that overlapped with the OC-CCI v2 temporal coverage  
138 (September 1997 - December 2013) were considered, and the concentrations of phytoplankton  
139 carbohydrate, protein and lipid were computed using the information on phytoplankton cell  
140 size (reported in Sal et al., 2013) and the corresponding allometric relationships (reported in  
141 Finkel et al., 2016a). This subset included 250 samples collected from 1997 to 2002, across  
142 various oceanographic regions (see Section 3.2, for the geographic locations); and consisted  
143 of 943 species of diatom, dinoflagellate and coccolithophores with equivalent-spherical diam-  
144 eter ranging from  $1.34\ \mu\text{m}$  to  $50\ \mu\text{m}$  (to be consistent with the size range of microplankton  
145 assumed within the algorithm, only the species with diameter  $<50\ \mu\text{m}$  were considered). This  
146 cell-diameter range covered nanoplankton, microplankton and a part of picoplankton. To be  
147 consistent with previous studies (Roy et al., 2013, 2017), the diameter ranges of the three  
148 phytoplankton size classes used in the model were picoplankton:  $0.2\text{--}2\ \mu\text{m}$ , nanoplankton:  
149  $2\text{--}20\ \mu\text{m}$ , and microplankton:  $20\text{--}50\ \mu\text{m}$ . Satellite matched-up chlorophyll concentrations and  
150 IOPs were retrieved from OC-CCI data archive. Given that the sampling times were mostly  
151 within the early years of SeaWiFS coverage (and SeaWiFS was the only contributing ocean-  
152 colour sensor over 1997-2002), a large number of gaps in satellite data were identified. To  
153 maximise the number of validation data points, match-ups from composite satellite images  
154 on daily ( $n = 39$ ) and monthly ( $n = 249$ ) scales were used.

155 The global annual stocks of phytoplankton carbohydrate, protein and lipid within the  
156 oceanic mixed layer were computed from the estimated surface concentrations, grid-by-grid,  
157 using the available mixed-layer depth values obtained from MIMOC (no specific depth profiles  
158 of the macromolecular concentrations were known from either *in situ* or remote sensing, on a  
159 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  $\text{pg cell}^{-1}$ ) and the volume of a phytoplankton cell ( $V_{cell}$ , in  $\mu\text{m}^3$ ) 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  $\xi$  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 ( $B_{total}$ ,  $\text{mg Chl m}^{-3}$ ) 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  $\text{mg m}^{-3}$ , 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:

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

Similarly, the total concentration of the macromolecule  $M$  (in  $\text{mg m}^{-3}$ ) 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  $[M]_{cell}$ :

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

with the condition that  $[M]_{total} \rightarrow \left[ 10^{-9} k a_M (10^{18} \pi/6)^{b_M} \log_e (D_{max}/D_{min}) \right]$ , 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  $\text{pg}$  to  $\text{mg}$ , and  $\text{m}^3$  to  $\mu\text{m}^3$  respectively. Using Eqs.(1) and (2), the ratio of the macromolecular concentration to the chlorophyll concentration ( $\chi_M$ ) can be expressed as:

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



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

$$M_{total} = \chi_M B_{total}, \quad (4)$$

193 provided that  $\xi$ ,  $a_M$  and  $b_M$  of the population are known (see Sections 2.3, 2.4).

## 194 2.3 Size-partitioned cellular contents of phytoplankton

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

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

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

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

202 and therefore,

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

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

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

204 Using the equations derived above, the concentrations of carbohydrate, protein and lipid  
 205 can be partitioned into any number of PSCs. However, for the sake of discussion, in this  
 206 study, the estimates are obtained for three major PSCs, namely, picoplankton, nanoplankton  
 207 and microplankton, with cell-diameter bounds  $[D_0, D_1]$ ,  $[D_1, D_2]$  and  $[D_2, D_3]$ , respectively,  
 208 where  $D_0 = 0.25 \mu\text{m}$ ,  $D_1 = 2 \mu\text{m}$ ,  $D_2 = 20 \mu\text{m}$ , and  $D_3 = 50 \mu\text{m}$  based on previous studies  
 209 (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 $\xi$ 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. (2016a); Hitchcock (1982); Menden-Deuer and Lessard (2000); Moal et al. (1987). More recently, Finkel et al. (2016a) 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. (2016a) (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  $\xi$  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  $\xi$ , 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

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

251 Nevertheless, to understand and quantify the overall uncertainty levels in the satellite-  
 252 derived estimates, a model sensitivity analysis was carried out. Theoretically, accurate esti-  
 253 mations of the macromolecular concentrations in phytoplankton based on the above method  
 254 would depend on the allometric parameters ( $a_M$  and  $b_M$ ) and the estimates of  $\xi$ . The re-  
 255 trieval of  $\xi$  further depends on satellite-derived estimates of chlorophyll-a and  $a_{ph}$ . Using  
 256 Eqs. (1-3), the relative sensitivities of the estimates of  $M_{total}$ , i.e.,  $\frac{\Delta M_{total}}{M_{total}}$ , can be computed  
 257 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  
 258 reported pixel-by-pixel in the global ocean, a maximum overall  $\frac{\Delta \xi}{\xi}$  in the range 0–25% is  
 259 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  
 260 reported by Finkel et al. (2016a) are considered. The resultant  $\frac{\Delta M_{total}}{M_{total}}$  are then computed  
 261 pixel-by-pixel, as percentages of the default estimates. So, without the availability of ade-  
 262 quate *in situ* measurements, the uncertainties discussed in the following sections should be  
 263 interpreted as model-based uncertainties, and not as those based on the *in situ* observations.

## 264 3 Results and discussion

### 265 3.1 Macromolecular concentrations across phytoplankton size range

266 The ratios of carbohydrate-to-chlorophyll ( $\chi_{carbo}$ ), protein-to-chlorophyll ( $\chi_{prot}$ ) and lipid-to-  
 267 chlorophyll ( $\chi_{lipid}$ ) increase with  $\xi$  within the ranges given by [5.0, 9.5], [7.1, 48.9] and [3.1,  
 268 32], respectively (Fig. 1a). For any given value of  $\xi$ ,  $\chi_{prot}$  is higher than  $\chi_{carbo}$  and  $\chi_{lipid}$ . For  
 269 low values of  $\xi$ ,  $\chi_{lipid}$  is lower than  $\chi_{carbo}$ , but it increases more rapidly with the assemblages  
 270 of small phytoplankton cells, and so, for high values of  $\xi$ ,  $\chi_{lipid}$  is significantly higher than  
 271  $\chi_{carbo}$  (Fig. 1a).

272 The proportions of carbohydrate, protein and lipid increase with  $\xi$  in picoplankton  
 273 (Fig. 1b), and decrease with  $\xi$  in microplankton (Fig. 1d), but are unimodal in nanoplank-  
 274 ton having magnitudes typically less than 50% with highest values in the middle range of  $\xi$   
 275 (Fig. 1c). At any given level of  $\xi$ , the proportion of lipid in picoplankton is higher than that of  
 276 carbohydrate or protein (with carbohydrate < protein < lipid) (Fig. 1b); but in microplank-  
 277 ton the order is reversed to carbohydrate > protein > lipid (Fig. 1d). For nanoplankton these  
 278 proportions alter from carbohydrate < protein < lipid at the lower end of  $\xi$  to carbohydrate >  
 279 protein > lipid at the higher end of  $\xi$  (Fig. 1c). These results show strong dependencies of phy-  
 280 toplankton size structure on the available macromolecular concentrations with implications  
 281 on their stocks in mixed populations of phytoplankton.

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

285 For protein estimates (Fig. 1f), the relative uncertainties would be <40% across the range of  
 286  $\xi$  provided that the relative uncertainty in  $\xi$  is <10%. If the relative uncertainties in  $\xi$  are  
 287 >15%, the uncertainties in protein would increase to >60% typically for  $3.25 < \xi < 4.5$ , but  
 288 would generally remain within <40% for populations dominated by either large or small cells  
 289 (i.e., at the low and high ends of  $\xi$ , see Table 1 for more details). For lipid estimates, the  
 290 relative uncertainties would be similar to those for protein: <40% for the low and high ends  
 291 of  $\xi$ , but >60% for the mid-range of  $\xi$ , if the uncertainties in  $\xi$  is >15% (Fig. 1g). Further  
 292 details on these uncertainty estimates for various combinations of uncertainties in  $\xi$  estimates  
 293 (based on Fig. 1e-g) are summarised in Table 1, and the propagations of the uncertainties in  
 294 the global ocean are discussed in Sections 3.7.

### 295 3.2 Comparison with estimates based on *in situ* abundance data

296 The matched-up *in situ* data were from specific cruises (see, Fig. 2a) with moderate sample  
 297 size having non-normal distribution; therefore, non-parametric statistics were implemented,  
 298 in particular, Spearman’s correlation instead of Pearson’s, and other non-parametric matrices  
 299 following Werdell et al. (2009). The *in situ* and satellite-based estimates generally follow  
 300 the 1 : 1 line, but with some level of spread around it (Fig. 2b-d, Supplementary Fig. S1),  
 301 with significant correlations (Spearman’s  $\rho$ ) between them on linear scale, for carbohydrate:  
 302  $\rho = 0.25$ ,  $p < 0.001$ ; protein:  $\rho = 0.24$ ,  $p < 0.001$ ; and lipid:  $\rho = 0.23$ ,  $p < 0.001$  (Fig. 2b-d).  
 303 The root mean squared error (RMSE) and bias of the estimates vary for carbohydrate (RMSE  
 304 10.20, bias  $-7.28 \text{ mg m}^{-3}$ ), protein (RMSE 21.55, bias  $-10.93 \text{ mg m}^{-3}$ ) and lipid (RMSE 9.77,  
 305 bias  $-4.87 \text{ mg m}^{-3}$ ). As expected, the RMSE and bias for daily match-ups, turn out to be  
 306 lower than those for monthly match-ups (see, Supplementary Table S1); but in both cases  
 307 their magnitudes are within a reasonable range, when compared with those for other derived  
 308 products, such as phytoplankton carbon (Kostadinov et al., 2016; Roy et al., 2017).

309 Following Werdell et al. (2009), three further metrics are computed for comparing the  
 310 estimates with daily (monthly) match-ups: the median satellite-to-in-situ-ratio (median ra-  
 311 tio, found to be 0.51(0.71), 0.59 (0.73), and 0.59 (0.73) respectively), the median of the  
 312 relative-percent difference (median RPD, found to be  $-49.41$  ( $-29.36$ ),  $-40.86$  ( $-27.38$ ) and  
 313  $-41.11$  ( $-26.75$ ), respectively), and the semi-interquartile percent differences (SIQ-PD, found  
 314 to be  $-48.50$  ( $-67.65$ ),  $-50.66$  ( $-65.08$ ) and  $-51.36$  ( $-63.82$ ), respectively) (see, Supplemen-  
 315 tary Table S1). The median RPDs and SIQ-PDs are lowest for lipid estimates, followed by  
 316 those for protein and carbohydrate (Supplementary Table S1). The median ratios are  $< 1$ ,  
 317 suggesting that the algorithm would generally underestimate the macromolecular concentra-  
 318 tions (Fig. 2e). Also, the algorithm seems to produce relatively less natural variability of the  
 319 macromolecular concentrations, in comparison with those estimated from the *in situ* abun-  
 320 dance data (Fig. 2e). However, it is worth mentioning that the median ratio, median RDP,  
 321 SIR-PD for SeaWiFS chlorophyll were reported (Werdell et al., 2009) to be in the ranges  
 322 [1.7, 81.5],  $[-34.7, 122.3]$ , and [0.88, 1.69], respectively. Therefore, in terms of these metri-  
 323 ces, the accuracy of the current estimates of the macromolecular concentrations are generally  
 324 comparable with that reported for SeaWiFS chlorophyll.

325 Nevertheless, these comparisons would be affected by several layers of uncertainties asso-  
 326 ciated with the *in situ* and satellite estimates. For example, prominent natural variability of  
 327 cell size of the 943 phytoplankton species would alter the *in situ* estimates of carbohydrate,  
 328 protein and lipid, which were not possible to include in the *in situ* calculations; and the un-  
 329 certainties in satellite inputs (chlorophyll, IOPs) would also affect the satellite retrievals of  $\xi$   
 330 (also see, Section 3.7).

### 331 **3.3 Phytoplankton carbohydrate, protein and lipid in the world** 332 **oceans**

333 Strong spatial variability of the annually-averaged  $\chi_{carbo}$ ,  $\chi_{prot}$ ,  $\chi_{lipid}$ , carbohydrate, protein  
 334 and lipid are found over the world's oceanic biomes, for the period of study (Fig. 3). The  
 335 magnitude of  $\chi_{carbo}$  varies from  $<5$  in the high-chlorophyll coastal waters and large parts of  
 336 the northern latitudes beyond 40 degree north (Fig. 3a,c), to  $>9$  in the open oceans and Case I  
 337 waters (Fig. 3c). Similarly,  $\chi_{prot}$  (Fig. 3e) or  $\chi_{lipid}$  (Fig. 3g) vary, respectively, from  $<15$  or  $<10$   
 338 in the coastal waters and northern latitudes, to  $>45$  or  $>30$ , respectively, in the open oceans  
 339 and Case I waters. These results generally reflect that the oceanographic regions dominated  
 340 by large and small phytoplankton are respectively represented by low and high values of  $\chi_{carbo}$ ,  
 341  $\chi_{prot}$  or  $\chi_{lipid}$ . In the Atlantic and Pacific subtropical gyres, despite the high magnitudes of  
 342  $\chi_{carbo}$ ,  $\chi_{prot}$  and  $\chi_{lipid}$ , the concentrations of carbohydrate, protein and lipid are typically low  
 343 ( $< 0.5$ ,  $1.0$  and  $1.0 \text{ mg m}^{-3}$ , respectively), and the spatial pattern is similar to the distribution  
 344 of low chlorophyll. Most of the coastal oceans and Case II waters are generally characterised  
 345 by higher than 5, 10 and  $10 \text{ mg m}^{-3}$  of carbohydrate, protein and lipid, respectively, which in  
 346 places spike beyond 50, 100 and  $100 \text{ mg m}^{-3}$ , respectively (Fig. 3d,f,h). It is noteworthy that  
 347 some of these very high values may be attributed to the uncertain or erroneous retrievals of  
 348 chlorophyll and other optical properties in the optically complex water (as also discussed in  
 349 Section 2.5).

350 Applying the macromolecular concentration-to-energy conversion factors, i.e,  $4.2 \text{ kcal g}^{-1}$   
 351 for carbohydrate,  $4.19 \text{ kcal g}^{-1}$  for protein,  $9.5 \text{ kcal g}^{-1}$  for lipid (Finkel et al., 2016a; Hitch-  
 352 cock, 1982), the chemical-energy values of the surface-ocean phytoplankton can be computed  
 353 (Fig. 3b,d,f,h). The annual average of the phytoplankton energy-value is generally less than  
 354  $0.1 \text{ kJ per m}^{-3}$  of ocean water in the subtropical gyres, but goes up to  $0.5\text{--}1.0 \text{ kJ per m}^{-3}$  in  
 355 parts of the equatorial, northern and southern latitudes, and beyond  $10 \text{ kJ per m}^{-3}$  in certain  
 356 coastal and optically complex waters (Fig. 3b).

### 357 **3.4 Size-partitioned phytoplankton carbohydrate, protein and lipid** 358 **in the world oceans**

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

362 In the northern latitudes beyond 40 degrees and in coastal waters, microplankton contribu-  
 363 tions to carbohydrate, protein and lipid are higher than those of picoplankton and nanoplank-  
 364 ton, with approximate ranges [2.5, 10], [2.0 25] and [0.5, 5.0]  $\text{mg m}^{-3}$ , respectively (Fig. 4c,f,i).  
 365 Nanoplankton contributions are generally in the range [1, 3]  $\text{mg m}^{-3}$  of carbohydrate, [1, 5]  
 366  $\text{mg m}^{-3}$  of protein and [1, 5]  $\text{mg m}^{-3}$  of lipid, respectively (Fig. 4b,e,h), except in the oligo-  
 367 trophic gyres, where all the concentrations reduce to less than  $0.05 \text{ mg m}^{-3}$  (Fig. 4b,e,h).

### 368 3.5 Macromolecular concentrations in Longhurst provinces

369 The geographical variations of carbohydrate, protein and lipid in the world oceans can be in-  
 370 ferred from their regionally-binned concentrations in the Longhurst biogeographical provinces  
 371 (Longhurst, 1995, 1998). Given that the ocean-colour data from satellites are inadequate (and  
 372 may be more erroneous) in the polar regions over most of the year, the estimates from the po-  
 373 lar provinces (6 out of 54 Longhurst provinces) are excluded from further discussion. For the  
 374 remaining 48 provinces, the spatial estimates of  $\chi_{carbo}$ ,  $\chi_{prot}$ ,  $\chi_{lipid}$  and the concentrations car-  
 375 bohydrate, protein and lipid are computed from their corresponding annually-averaged global  
 376 maps (Fig. 5). These provinces include 14 Westerlies (NADR, GFST, NASW, MEDI, NASE,  
 377 PSAE, PSAW, KURO, NPPF, NPSW, TASM, SPSG, SSTC, SANT), 12 Trades (NATR,  
 378 WTRA, ETRA, SATL, CARB, MONS, ISSG, NPTG, PNEC, PEQD, WARM, ARCH) and  
 379 22 Coastal (NECS, CNRY, GUIN, GUIA, NWCS, BRAZ, FKLD, BENG, EAFR, REDS,  
 380 ARAB, INDE, INDW, AUSW, ALSK, CCAL, CAMR, CHIL, CHIN, SUND, AUSE, NEWZ)  
 381 provinces (full names of the provinces are given in Supplementary Table S2, and the descrip-  
 382 tions in Longhurst, 1995, 1998). The Westerlies, Trades and Coastal provinces are shown in  
 383 Fig. 2a.

384 Spatial variability of the estimates in the Coastal provinces are found to be higher than  
 385 those in the Westerlies or Trades provinces (Fig. 5), with the lowest variability in the West-  
 386 erlies provinces (Fig. 3-5), reflecting that coastal upwellings would strongly influence the dis-  
 387 tribution of phytoplankton macromolecules (similar to chlorophyll distribution). The spatial  
 388 medians of  $\chi_{carbo}$ ,  $\chi_{prot}$  and  $\chi_{lipid}$  are lowest (5.69, 13.86 and 8.0, respectively) for the NWCS  
 389 (North-West Atlantic Coastal Shelves) province, and highest (8.95, 45.13, 29.56, respectively)  
 390 for the NPSW (North Pacific Subtropical Gyre West) province (Fig. 5a,b,c, and Supplemen-  
 391 tary Table S2). The NECS (North-East Atlantic Coastal Shelves) province is characterised  
 392 by the highest surface concentrations (Fig. 5d,e,f) of the annually-averaged spatial medians  
 393 of carbohydrate ( $9.53 \text{ mg m}^{-3}$ ), protein ( $25.2 \text{ mg m}^{-3}$ ), and lipid ( $14.81 \text{ mg m}^{-3}$ ). The low-  
 394 est surface concentrations (spatial median) of carbohydrate ( $0.60 \text{ mg m}^{-3}$ ) and lipid ( $1.75$   
 395  $\text{mg m}^{-3}$ ) are obtained in the NATR (North Atlantic Tropical Gyral) province (Fig. 5d,f, Ta-  
 396 ble S2), whereas, the lowest concentrations of protein ( $2.11 \text{ mg m}^{-3}$ ) is obtained in the NPSW  
 397 (North Pacific Subtropical Gyre West) province (Fig. 5e Table S2), both of which are generally  
 398 populated by small picoplankton throughout the year.

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

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

### 414 3.6 Global-ocean stocks of phytoplankton macromolecules

415 The annually-averaged global stocks are: 0.044 Gt of carbohydrate with monthly range [0.041,  
416 0.05] Gt; 0.17 Gt of protein with monthly range [0.155, 0.18] Gt; and 0.108 Gt of lipid with  
417 monthly range [0.098, 0.121] Gt (Fig. 6, and Supplementary Table S4). The largest global  
418 stocks are obtained in the month of September, which generally matches with the time of  
419 phytoplankton bloom in large parts of the equatorial-southern hemisphere (Kostadinov et al.,  
420 2017). The smallest stocks are obtained in the month of June, generally after the termination  
421 of the spring blooms.

422 The percentages of the size-partitioned carbohydrate, protein and lipid stocks also vary  
423 over the months of the years (Fig. 6). The stocks constitute the lowest percentage of picoplank-  
424 ton carbohydrate ~46% (equivalent to 0.02 Gt, with monthly range of 43-53%), compared with  
425 the percentages of picoplankton protein ~78% (equivalent to 0.133 Gt, with monthly range of  
426 76-83%), and picoplankton lipid ~85% (equivalent to 0.092 Gt, with monthly range of 83-88%)  
427 (Supplementary Table S4). The stocks further constitute ~33% of nanoplankton carbohydrate  
428 (equivalent to 0.015 Gt, with monthly range of 32-36%), which is considerably higher than the  
429 percentages of nanoplankton protein ~17% (equivalent to 0.028 Gt, with monthly range of  
430 14-18%), and nanoplankton lipid ~12% (equivalent to 0.013 Gt, with monthly range of 10-  
431 13%). Similarly, the percentage of microplankton carbohydrate ~21% (equivalent to 0.009 Gt,  
432 with monthly range of 16-24%) is significantly higher than the percentages of microplankton  
433 protein ~5% (equivalent to 0.009 Gt, with monthly range of 3-7%) and microplankton lipid  
434 ~3% (equivalent to 0.003 Gt, with monthly range of 2-4%). But clearly, for any given macro-  
435 molecular stock, the largest contribution comes from picoplankton and the smallest from  
436 microplankton (Fig. 6).

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

440 estimated previously from satellite remote sensing (e.g., Behrenfeld et al., 2005; Kostadinov  
441 et al., 2016; Roy et al., 2017). For example, the annually-averaged stocks of the total phyto-  
442 plankton biomass varied between 0.2 GtC to 1.0 GtC depending on the estimation method  
443 (e.g., Behrenfeld et al., 2005; Falkowski et al., 1998; Kostadinov et al., 2016; Roy et al., 2017;  
444 Stramski et al., 2008; Taylor et al., 2012). The annually-averaged stocks of carbohydrate  
445 protein and lipid and their sum total, estimated above, are within this range. Recent studies  
446 (Finkel et al., 2016*a,b*) also suggested that under ‘nutrient-sufficient, exponential growth con-  
447 ditions’ the median composition of the dry weight of microalgae contains 15% carbohydrate  
448 32.2% protein and 17.3% lipid. With respect to the most recent satellite-based estimates of  
449 phytoplankton biomass (i.e.,  $\sim 0.3$  GtC, based on Kostadinov et al., 2016; Roy et al., 2017),  
450 the percentages of the annually-averaged global stocks (which included both nutrient suffi-  
451 cient and oligotrophic waters) of carbohydrate, protein and lipid are  $\sim 15\%$ ,  $\sim 57\%$ ,  $\sim 36\%$ ,  
452 respectively. These preliminary results thus suggests that on a global scale, the relative pro-  
453 portions of carbohydrate in phytoplankton might be more robust than the proportions of  
454 protein and lipid. However, direct *in situ* measurements would be required to further validate  
455 these results.

### 456 **3.7 Algorithm uncertainties on global map**

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

## 470 **4 Concluding remarks**

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



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

488 Recent studies based on either ocean-colour data (Behrenfeld et al., 2005; Kostadinov et al.,  
489 2016; Roy et al., 2017; Stramski et al., 2008) or Earth System models (e.g., CMIP5, Taylor  
490 et al., 2012), have attempted to improve the estimates of the stocks of phytoplankton carbon,  
491 and have narrowed down the estimation range of the annually-averaged stocks. But unclear is  
492 how the total carbon stock partitions into the stocks of essential carbon-based macromolecules  
493 in phytoplankton. For example, although the proportions of the macromolecules to dry weight  
494 of phytoplankton are reported for ideal nutrient-rich conditions (Finkel et al., 2016*a,b*), little  
495 is known about those proportions in diverse oceanographic regions where growth conditions  
496 deviate from ideal. This study independently estimates the annually-averaged stocks of the  
497 three essential phytoplankton macromolecules, and finds that the sum total of these estimates  
498 are well within the range of the reported stocks of total phytoplankton carbon. The estimates  
499 would be potentially useful for understanding the cellular allocation of carbon to carbohydrate,  
500 protein and lipid pools in phytoplankton, both spatially and over time, with implications for  
501 trophic transfer models, and higher trophic or fisheries models.

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

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

## 524 Acknowledgements

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

## 533 References

- 534 Behrenfeld, M. J., Boss, E., Siegel, D. A. and Shea, D. M. (2005), 'Carbon-based ocean  
535 productivity and phytoplankton physiology from space', *Global biogeochemical cycles* **19**(1),  
536 GB1006.
- 537 Breteler, W. K., Schogt, N. and Rampen, S. (2005), 'Effect of diatom nutrient limitation on  
538 copepod development: role of essential lipids', *Marine Ecology Progress Series* **291**, 125–  
539 133.
- 540 Brewin, R. J., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E.  
541 et al. (2015), 'The ocean colour climate change initiative: Iii. a round-robin comparison on  
542 in-water bio-optical algorithms', *Remote Sensing of Environment* **162**, 271–294.
- 543 Budge, S. M., Devred, E., Forget, M.-H., Stuart, V., Trzcinski, M. K., Sathyendranath, S.  
544 et al. (2014), 'Estimating concentrations of essential omega-3 fatty acids in the ocean:  
545 supply and demand', *ICES Journal of Marine Science: Journal du Conseil* **71**(7), 1885–  
546 1893.
- 547 Chisti, Y. (2007), 'Biodiesel from microalgae', *Biotechnology advances* **25**(3), 294–306.
- 548 Falkowski, P. (2012), 'The power of plankton', *Nature* **483**, S17–S20.
- 549 Falkowski, P. G., Barber, R. T. and Smetacek, V. (1998), 'Biogeochemical controls and feed-  
550 backs on ocean primary production', *Science* **281**(5374), 200–206.

- 551 Field, C. B., Behrenfeld, M. J., Randerson, J. T. and Falkowski, P. (1998), ‘Primary  
552 production of the biosphere: integrating terrestrial and oceanic components’, *Science*  
553 **281**(5374), 237–240.
- 554 Finkel, Z., Follows, M. and Irwin, A. (2016a), ‘Size-scaling of macromolecules and chemical  
555 energy content in the eukaryotic microalgae’, *Journal of Plankton Research* **38**(5), 1151–  
556 1162.
- 557 Finkel, Z., Follows, M. J., Liefer, J. D., Brown, C. M., Benner, I. and Irwin, A. J.  
558 (2016b), ‘Phylogenetic diversity in the macromolecular composition of microalgae’, *PloS*  
559 *one* **11**(5), e0155977.
- 560 GCOS, G. (2011), ‘Systematic observation requirements for satellite-based products for cli-  
561 mate. 2011 update supplement details to the satellite 39 based component of the imple-  
562 mentation plan for the global observing system for climate in support of the unfccc (2010  
563 update)’, Tech. rep., World Meteorological Organisation (WMO), 7 bis, avenue de la Paix,  
564 CH- 1211 Geneva 2, Switzerland.
- 565 Geider, R. and La Roche, J. (2002), ‘Redfield revisited: variability of c: N: P in marine  
566 microalgae and its biochemical basis’, *European Journal of Phycology* **37**(1), 1–17.
- 567 Goldman, J. C. and Caron, D. A. (1985), ‘Experimental studies on an omnivorous microflagel-  
568 late: implications for grazing and nutrient regeneration in the marine microbial food chain’,  
569 *Deep Sea Research Part A. Oceanographic Research Papers* **8**, 899–915.
- 570 Hitchcock, G. L. (1982), ‘A comparative study of the size-dependent organic composition of  
571 marine diatoms and dinoflagellates’, *Journal of plankton research* **4**(2), 363–377.
- 572 IOCCG (2000), Remote sensing of ocean colour in coastal, and other optically-complex, wa-  
573 ters, Technical Report 3, Dartmouth, NS.
- 574 IOCCG (2006), Remote sensing of inherent optical properties: Fundamentals, tests of algo-  
575 rithms, and applications, Technical Report 5, Dartmouth, Canada.
- 576 IOCCG, ed. (2014), *Phytoplankton functional types from Space.*, number 15, International  
577 Ocean-Colour Coordinating Group, Reports of the International Ocean-Colour Coordinat-  
578 ing Group (IOCCG).
- 579 Jónasdóttir, S. (1994), ‘Effects of food quality on the reproductive success of *acartia tonsa*  
580 and *acartia hudsonica*: laboratory observations’, *Marine Biology* **121**(1), 67–81.
- 581 Kostadinov, T., Milutinovic, S., Marinov, I. and Cabré, A. (2016), ‘Carbon-based phytoplank-  
582 ton size classes retrieved via ocean color estimates of the particle size distribution’, *Ocean*  
583 *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., Rodriguez, 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. (2007a), ‘Enrichment and ecosystem stability: effect of toxic  
635 food’, *BioSystems* **90**(1), 151–160.
- 636 Roy, S. and Chattopadhyay, J. (2007b), ‘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.

- 652 Schmidtko, S., Johnson, G. C. and Lyman, J. M. (2013), ‘MIMOC: A global monthly isopyc-  
 653 nal upper-ocean climatology with mixed layers’, *Journal of Geophysical Research: Oceans*  
 654 **118**(4), 1658–1672.
- 655 Sheldon, R., Prakash, A. and Sutcliffe, W. (1972), ‘The size distribution of particles in the  
 656 ocean’, *Limnology and oceanography* **17**(3), 327–340.
- 657 Shin, K., Jang, M.-C., Jang, P.-K., Ju, S.-J., Lee, T.-K. and Chang, M. (2003), ‘Influence  
 658 of food quality on egg production and viability of the marine planktonic copepod *acartia*  
 659 *omorii*’, *Progress in Oceanography* **57**(3), 265–277.
- 660 Sieburth, J. M., Smetacek, V. and Lenz, J. (1978), ‘Pelagic ecosystem structure: Het-  
 661 erotrophic compartments of the plankton and their relationship to plankton size fractions  
 662 1’, *Limnology and Oceanography* **23**(6), 1256–1263.
- 663 Sterner, R. W. and Elser, J. J. (2002), *Ecological stoichiometry: the biology of elements from*  
 664 *molecules to the biosphere*, Princeton University Press.
- 665 Stramski, D., Reynolds, R. A., Babin, M., Kaczmarek, S., Lewis, M. R., Röttgers, R. et al.  
 666 (2008), ‘Relationships between the surface concentration of particulate organic carbon and  
 667 optical properties in the eastern south pacific and eastern atlantic oceans’, *Biogeosciences*  
 668 **5**(1), 171–201.
- 669 Strathmann, R. (1967), ‘Estimating the organic carbon content of phytoplankton from cell  
 670 volume or plasma volume’, *Limnology and Oceanography* **12**(3), 411–418.
- 671 Taylor, K. E., Stouffer, R. J. and Meehl, G. A. (2012), ‘An overview of cmip5 and the  
 672 experiment design’, *Bulletin of the American Meteorological Society* **93**(4), 485–498.
- 673 Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A. and Marty, J.-C. (2001), ‘Phyto-  
 674 plankton pigment distribution in relation to upper thermocline circulation in the eastern  
 675 mediterranean sea during winter’, *Journal of Geophysical Research: Oceans (1978–2012)*  
 676 **106**(C9), 19939–19956.
- 677 Werdell, P. J., Bailey, S. W., Franz, B. A., Harding, L. W., Feldman, G. C. and McClain, C. R.  
 678 (2009), ‘Regional and seasonal variability of chlorophyll-a in chesapeake bay as observed  
 679 by seawifs and modis-aqua’, *Remote Sensing of Environment* **113**(6), 1319–1330.1

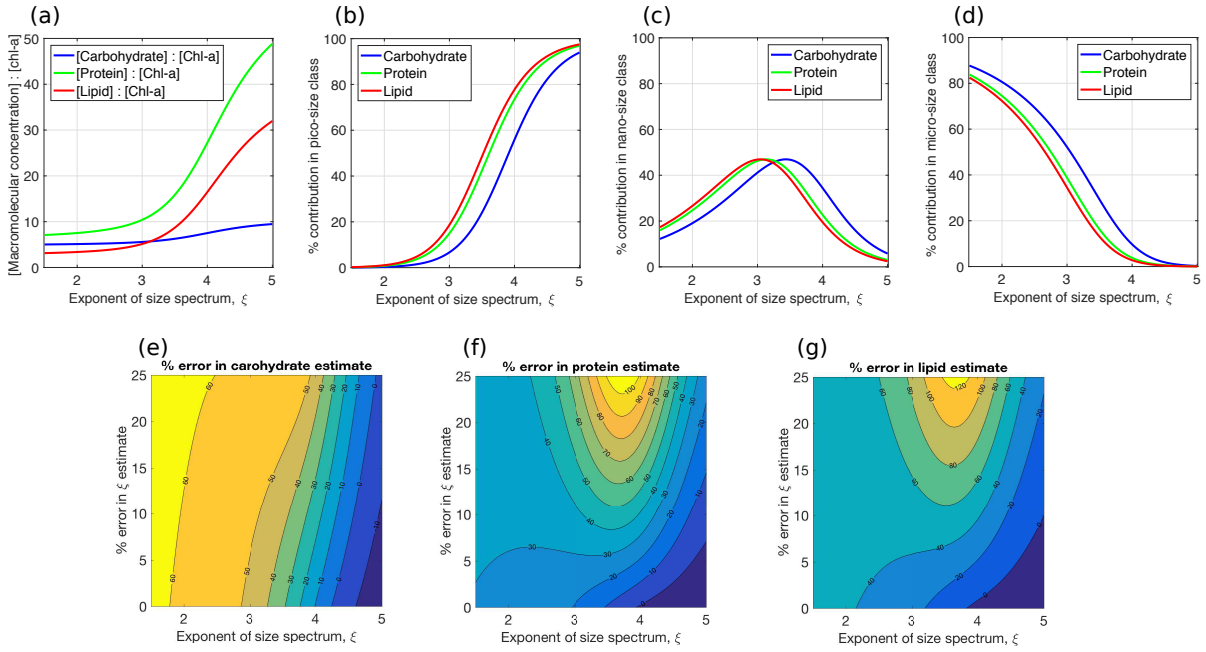
680 **Figure captions**


Figure 1: (a) Carbohydrate-to-chlorophyll ( $\chi_{carbo}$ ), protein-to-chlorophyll ( $\chi_{prot}$ ) and lipid-to-chlorophyll ( $\chi_{lipid}$ ) ratios of the mixed-phytoplankton population derived (using Eq. 3) as functions of the exponent of the phytoplankton size spectrum ( $\xi$ ). (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  $\xi$ ,  $a_M$  and  $b_M$  (see, Section 2.5). The 95% confidence levels for the allometric parameters reported in Finkel et al. (2016a) 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  $\xi$  (following Roy et al. (2013)).

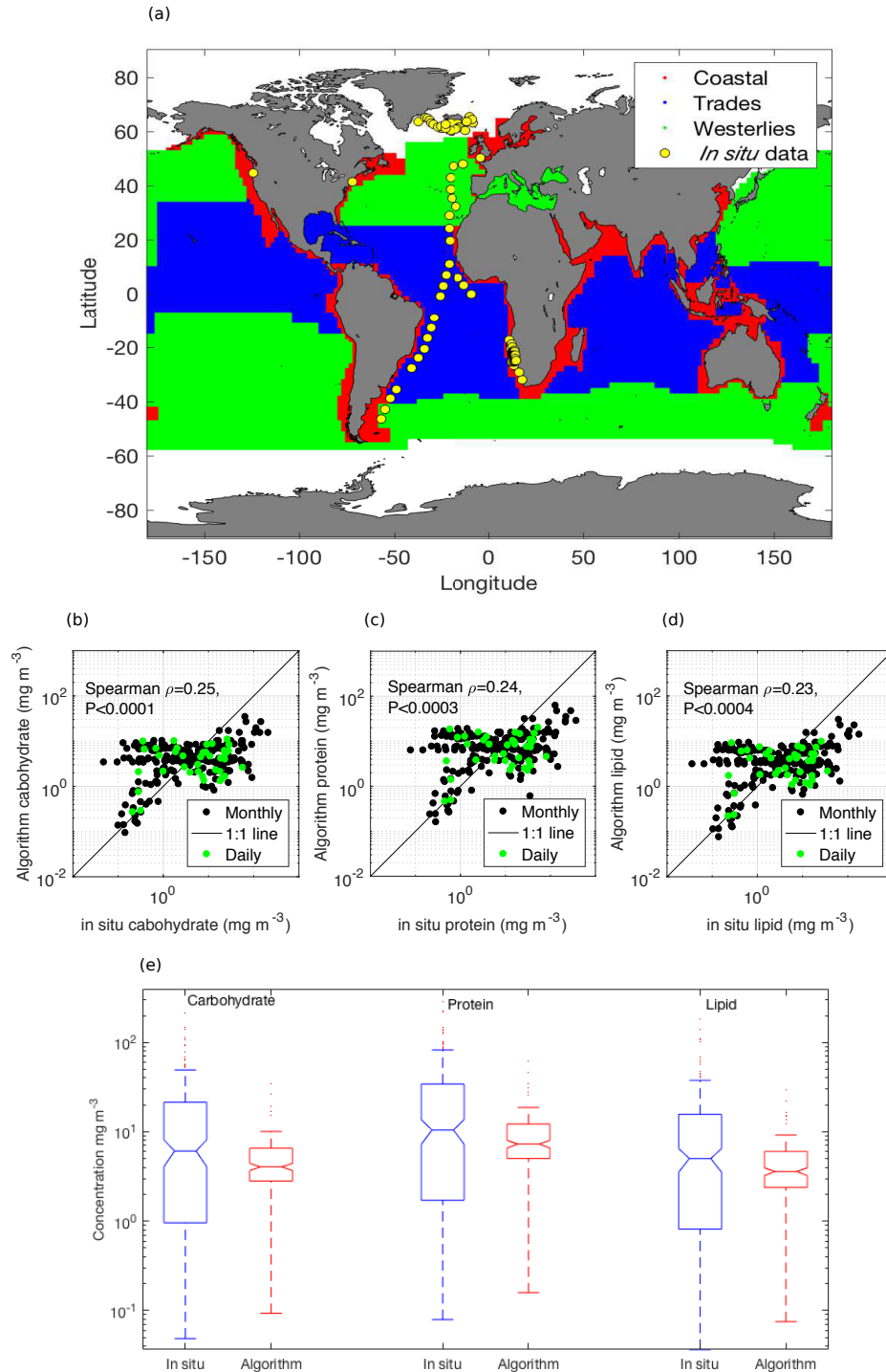


Figure 2: (a) Geographic locations of the *in situ* samples (yellow dots) used from the marine biodiversity database (Sal et al., 2013); this subset overlapped with the temporal coverage of satellite data, and were considered for computing phytoplankton carbohydrate, protein and lipid using species size and cell abundances, and by applying the allometric relationships reported in Finkel et al. (2016a). 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.



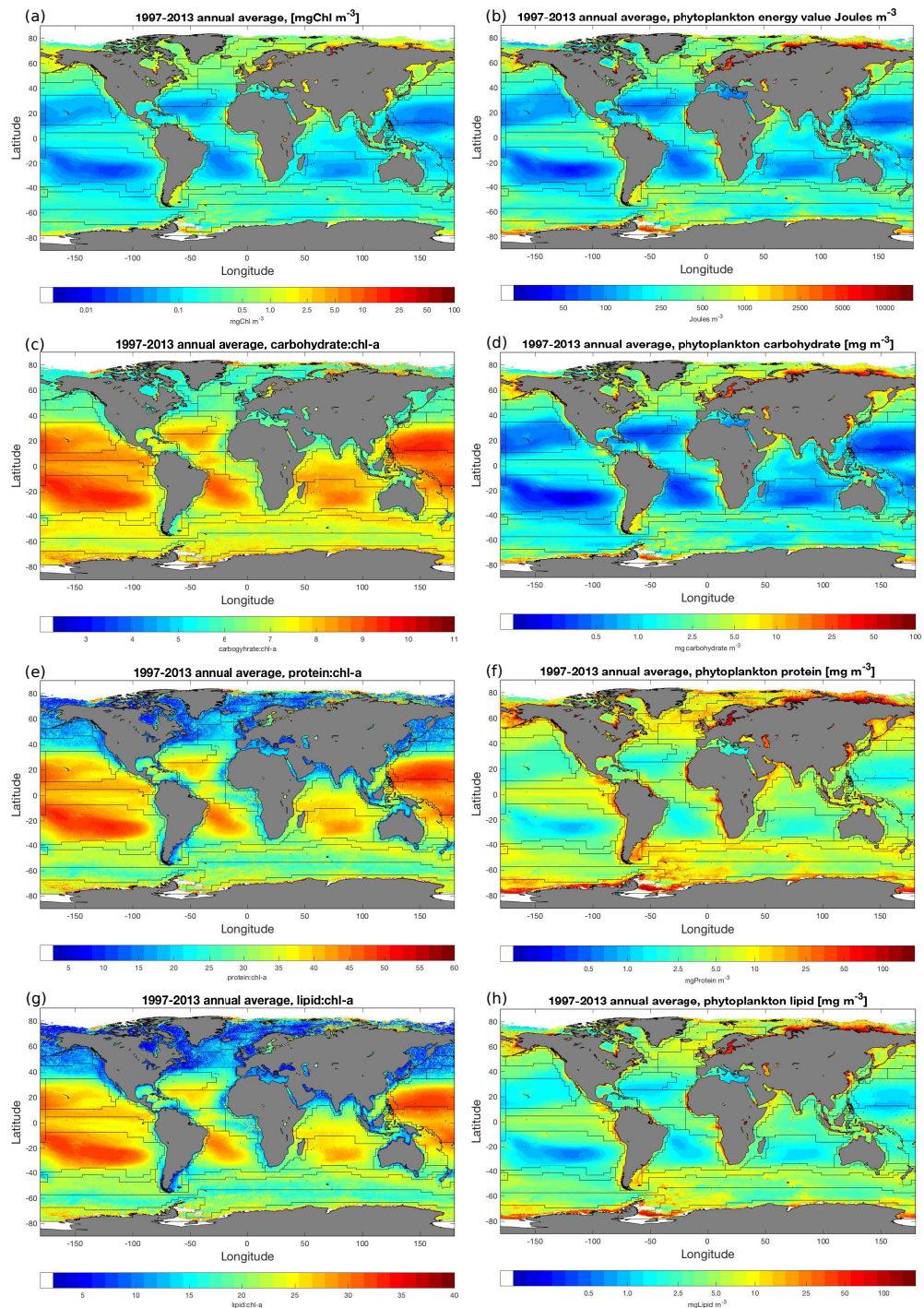


Figure 3: Distributions of the annually-averaged surface concentrations of macromolecules and energy value of phytoplankton over 1997-2013. Overlaid on the global maps are thin black lines representing the boundaries of the Longhurst biogeographical provinces (Longhurst, 1995, 1998). Annual averages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

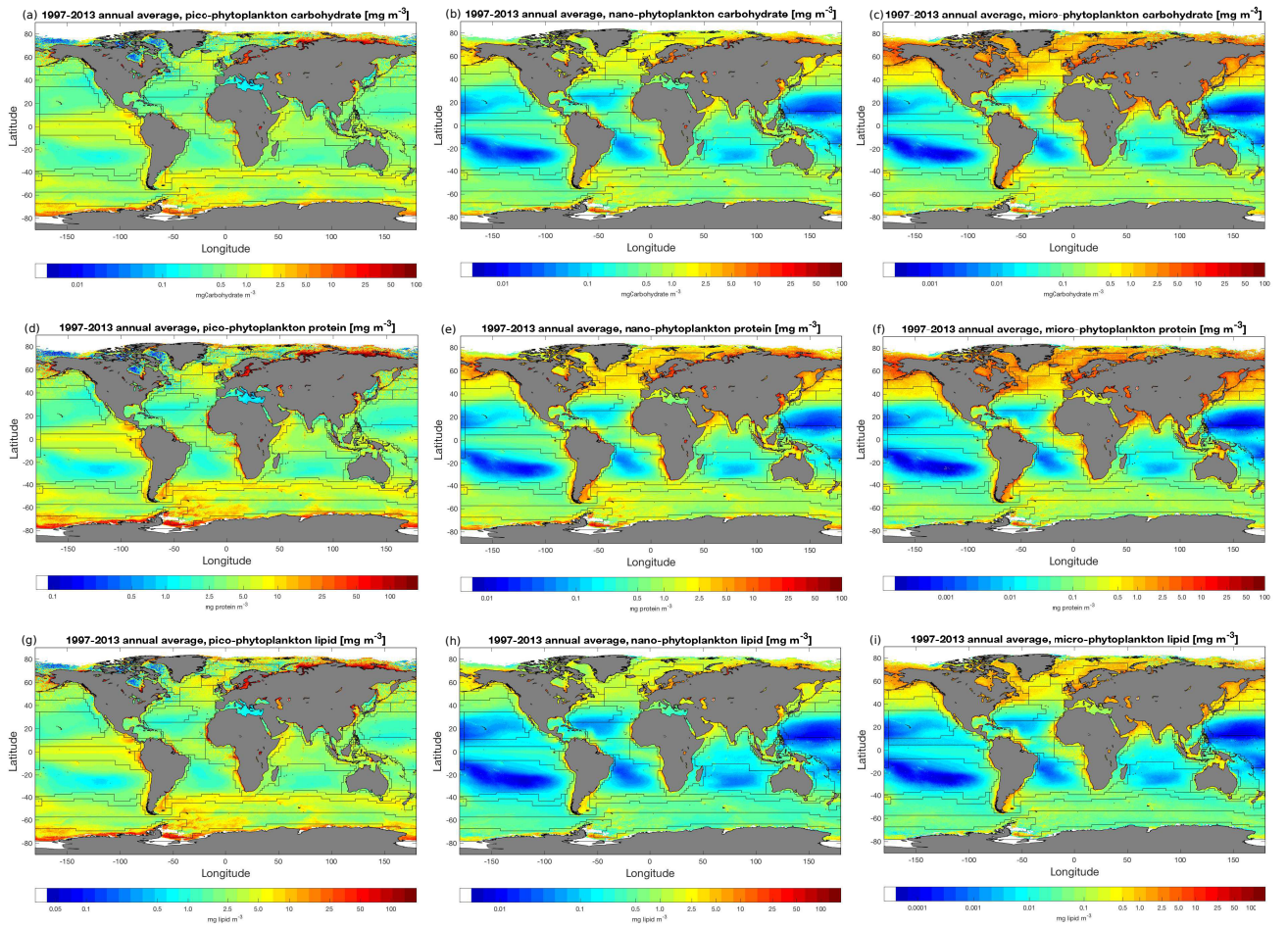


Figure 4: Annually-averaged surface macromolecular concentrations [ $\text{mg m}^{-3}$ ] in picoplankton, nanoplankton, and microplankton over 1997-2013: (a) picoplankton carbohydrate, (b) nanoplankton carbohydrate, (c) microplankton carbohydrate; (d) picoplankton protein, (e) nanoplankton protein, (f) 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).

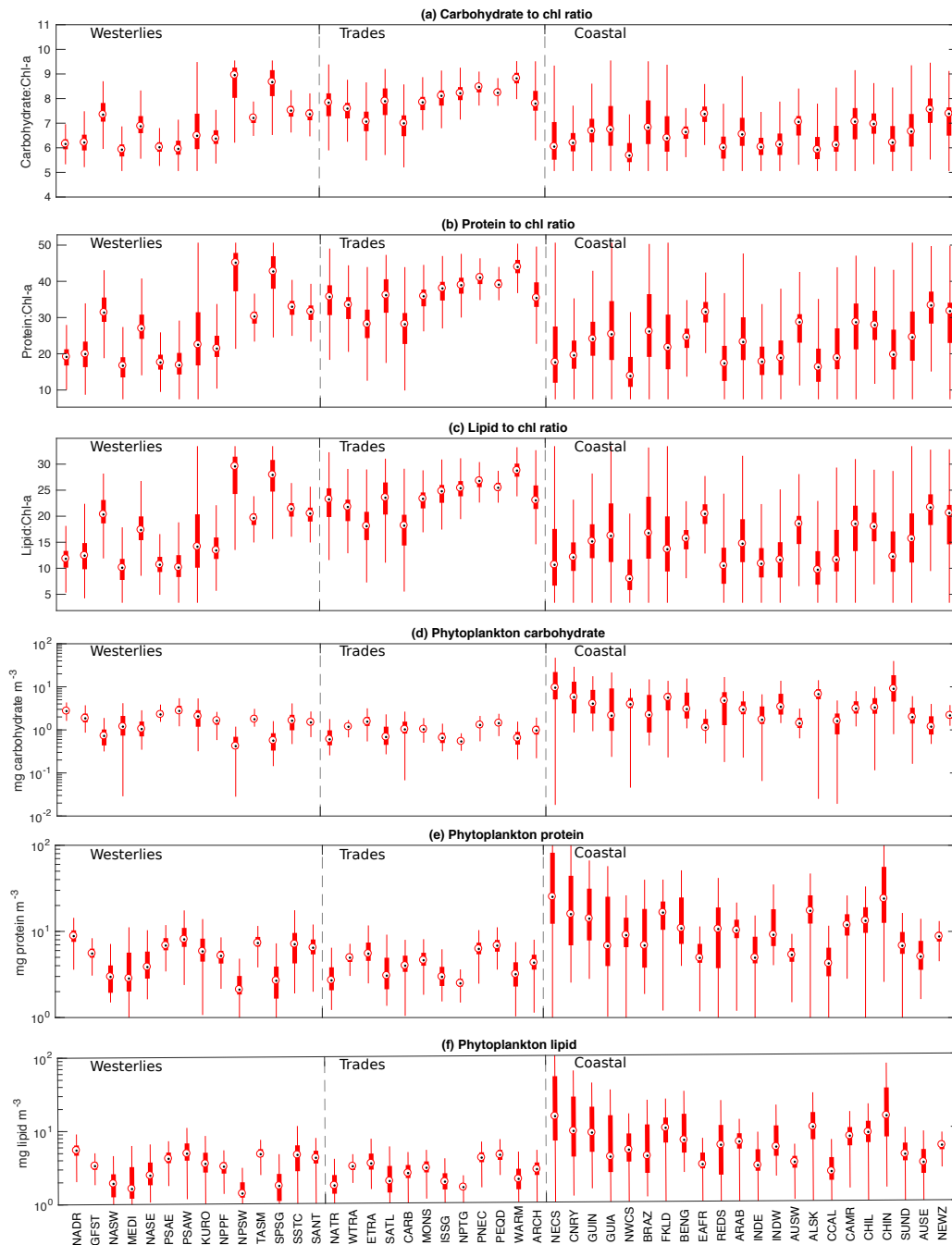


Figure 5: Annually-averaged surface macromolecular composition within Longhurst biogeographical provinces (Longhurst, 1995) computed over 1997-2013. Box plots with annual median (black dots), interquartile ranges (thick red bar), and ranges (thin whiskers) for (a) carbohydrate-to-chlorophyll ratio ( $\chi_{carbo}$ ), (b) protein-to-chlorophyll ratio ( $\chi_{prot}$ ), (c) lipid-to-chlorophyll ratio ( $\chi_{lipid}$ ), (d) carbohydrate ( $\text{mg m}^{-3}$ ), (e) protein ( $\text{mg m}^{-3}$ ), and (f) lipid ( $\text{mg m}^{-3}$ ), 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, GUID, 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.

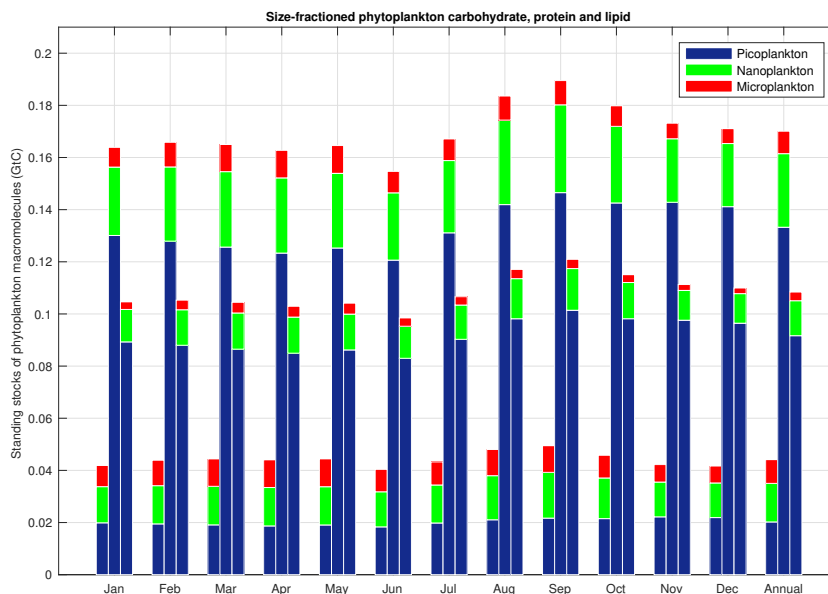


Figure 6: Annually-averaged macromolecular compositions for three phytoplankton size classes over 1997-2013. Grouped bars represent the monthly and annual stocks of the total (height of each bar) and size-partitioned (blue - picoplankton fraction, green - nanoplankton fraction, and red - microplankton 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).

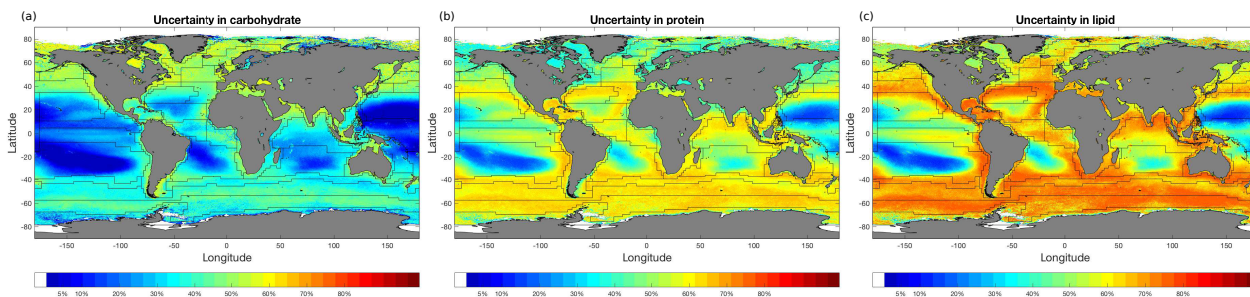


Figure 7: Algorithm uncertainty maps corresponding to the estimates of (a) phytoplankton carbohydrate, (b) phytoplankton protein and (c) phytoplankton lipid based on the sensitivity analysis in Section 2.5. Annually-averaged uncertainties in estimating the surface concentrations of carbohydrate, protein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

Table 1: Summary of overall uncertainties (mean with ranges) in carbohydrate, protein and lipid estimates as a function of uncertainties in  $\xi$  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