Scratching beneath the surface: a model to predict the vertical distribution of *Prochlorococcus* using remote sensing

Priscila K. Lange[1*], Robert J.W. Brewin[2,3], Giorgio Dall'Olmo[2,3], Glen A. Tarran[2], Shubha Sathyendranath[2,3], Mikhail Zubkov[4], Heather A. Bouman[1]

[1] Department of Earth Sciences, University of Oxford, South Parks Rd, Oxford, OX1 3AN, United Kingdom
[2] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, United Kingdom
[3] National Centre for Earth Observation, Plymouth Marine Laboratory, Plymouth PL1 3DH, United Kingdom
[4] National Oceanography Centre Southampton, Waterfront Campus, Southampton, SO14 3ZH, United Kingdom
* corresponding author: prilange@gmail.com

Supplementary Material

1. In-situ data

Prochlorococcus cell counts were determined by fixing 0.8 ml water samples using paraformaldehyde, and quantifying cells in these samples using a FACSCalibur (Becton Dickinson) flow cytometer. On the Atlantic Meridional Transect (AMT) cruises 12, 13, 16, 18, 19, 20, and 21, *Prochlorococcus* populations were determined based on their autofluorescence and right-angle light scatter (RALS) in unstained samples analysed following the method described in Heywood et al. (2006). On AMTs 14, 15, 17, 22, 23, and 24, samples were stained with 1% commercial stock solution of SYBR Green 1 (Molecular Probes, Inc.) in Milli-Q water mixed with 300 mM tripotassium citrate (24.5 mM final concentration) immediately following fixation (Olson et al., 1993; Zubkov et al., 2007). *Prochlorococcus* populations were determined using the cells' green fluorescence and right-angle light scatter, and yellow-green 0.5 and 1.0 μm reference beads (Fluoresbrite Microparticles, Polysciences, Warrington, PA, USA) were used as an internal standard for both fluorescence and flow rates (Zubkov and Burkill, 2006). The WinMDI (version 2.0) freeware (Joseph Trotter) was used to extract data of the flow cytometer output.

In-situ Z_{DCM} (Z_{DCM}') was defined as the depth at which the peak in chlorophyll (red) fluorescence was observed in each profile, using a SeaBird fluorometer mounted on a CTD rosette. The *in-situ* attenuation coefficient for PAR (K_dPAR') was calculated from profiles of downwelling PAR measured using a SeaBird CTD PAR sensor, then K_dPAR' was used to calculate the *in-situ* fractional PAR at depth ($fPAR(z)'$), or the percent value of surface irradiance at depth. Remote sensing reflectance at 443 and 488 nm were acquired from match-ups with 4×4 km 8-day satellite-derived composites (Aqua-MODIS, Ocean Color, NASA 2014).

The compilation, processing, and analysis of the *in-situ* dataset were performed using R packages *stats* (R Core Team, 2017), *ggplot2* (Wickham, 2009), *lattice* (Sarkar, 2008) and *devtools* (Wickham

and Chang, 2016). Satellite data were assessed using the R packages *ncdf4* (Pierce, 2015), *chron* (James and Hornik, 2017), *fields* (Nychka et al., 2015), and *raster* (Hijmans, 2016).

2. Delineation of the area of subtropical gyres

Subtropical gyres are vast oligotrophic areas usually in the centre of ocean basins. These areas present very low chlorophyll concentrations at the sea surface due to the limitation of nutrients, as they are remote from any nutrient source and, usually, the pycnocline is deep and strong, slowing the transport of nutrients from depth to the euphotic zone.

As the depth of the nutricline commonly coincides with the depth of the deep chlorophyll maximum (Z_{DCM}) (**Fig. S1**), Z_{DCM} can be used as a proxy for nutrient supply from depth in these areas.

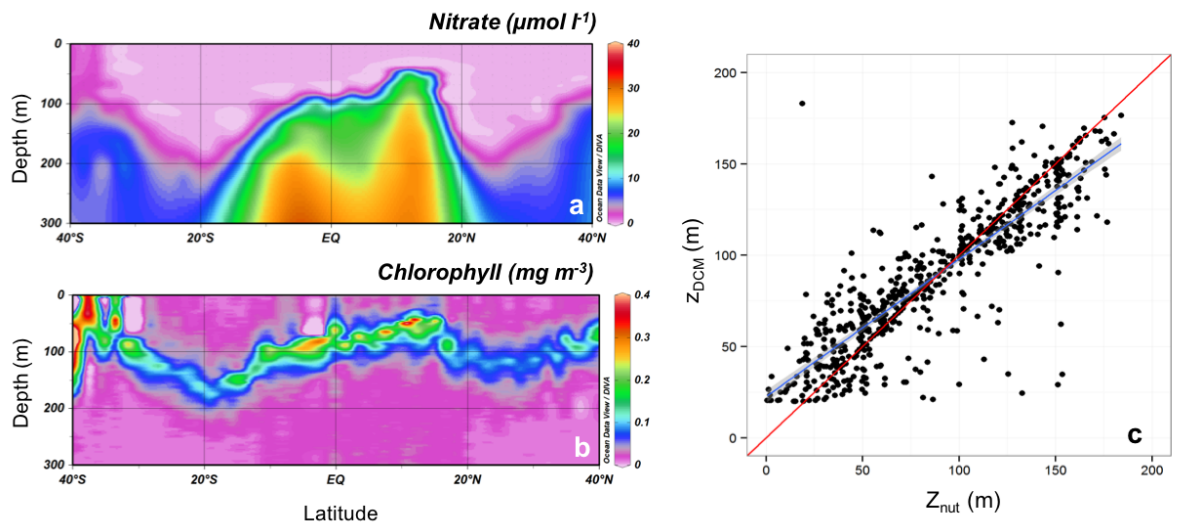

Figure S1. Latitudinal section of the Atlantic Ocean, showing profiles of a) the concentration of nitrate and b) the concentration of chlorophyll. Figure c) shows the relationship between the depth of the nutricline (Z_{nut}) and the depth of the deep chlorophyll maximum (Z_{DCM}), where the red line shows the 1:1 relationship. Data for figs. (a) and (b) from AMT 18 (1998), and data for fig. (c) from AMTs 12 to 24 (n=533 observations), were provided by the British Oceanographic Data Centre (BODC).

In **Fig. S1c**, Z_{nut} was defined as the depth where the concentration of nitrate becomes 0.5 µmol l^{-1} higher than the concentration of nitrate at surface, and Z_{DCM} is the depth where the fluorescence-derived chlorophyll concentration is highest (when this depth > 50 m).

Some criteria are commonly used to delineate the ocean gyres, such as:

a) sea surface chlorophyll ≥ 0.07 mg m^{-3} (McClain et al., 2004);

b) the depth of the deep chlorophyll maximum (Z_{DCM}) ≥ 80 metres (Zubkov et al., 2000); and

c) the temperature at 100 metres depth ≥ 17°C (Zubkov et al., 2000).

Considering the relationship between Z_{DCM} and sea-surface chlorophyll (*SSChlo*) found in our dataset of 12 AMT cruises (**Fig. S2**), we established a *SSChlo* threshold of 0.075 mg m^{-3} to delineate the area of the ocean gyres.

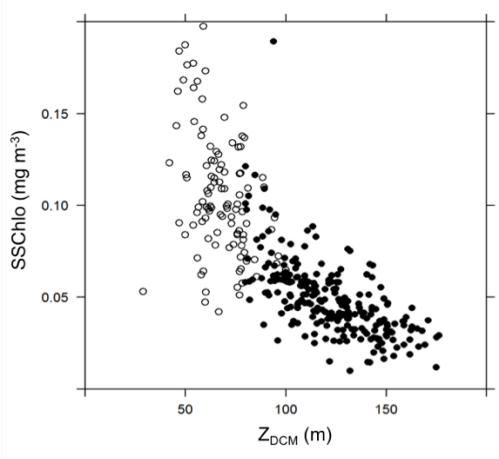

Figure S2. Relationship between the depth of the deep chlorophyll maximum (Z_{DCM}) and the sea-surface chlorophyll (*SSChlo*) (n=409 observations). Black dots show locations where Z_{DCM} >= 80 metres. Data from AMTs 12 to 24 were (provided by the British Oceanographic Data Centre (BODC).

3. Equations to estimate the vertical distribution of light

The following equations were used to estimate the vertical distribution of light in the water column. The *in-situ* attenuation coefficient for PAR (K_dPAR') was calculated from profiles of downwelling PAR according to equation (S1) from Kirk (2011):

$$K_dPAR' = -d \ln E_d / dz \,, \tag{S1}$$

where E_d is the downwelling irradiance, and z is the depth in metres. Then, K_dPAR' was used to calculate the *in-situ* fractional PAR at depth ($fPAR(z)'$) in equation (S2) from Kirk (2011):

$$fPAR(z)' = \exp(-K_dPAR'\ z) \,. \tag{S2}$$

The equation used to calculate the satellite-derived K_dPAR (equation (1) from the manuscript) was derived from K_dPAR'. The rationale for using equation (1) rather than equation (9') from Morel et al. (2007) that is conventionally used to estimate K_dPAR under open ocean conditions is that, in clear oligotrophic waters, *Prochlorococcus* dominates. We would argue that to achieve the best estimates of the *Prochlorococcus* standing stock, which dominates vast regions of the global ocean that are characterised by low chlorophyll and consequently low K_dPAR, it is critical to have accurate estimates of K_dPAR when light attenuation is low. When comparing equation (1) to K_dPAR values estimated using Morel et al.'s equation, the estimates of equation (1) were more similar to *in-situ* values (**Fig. S3**).

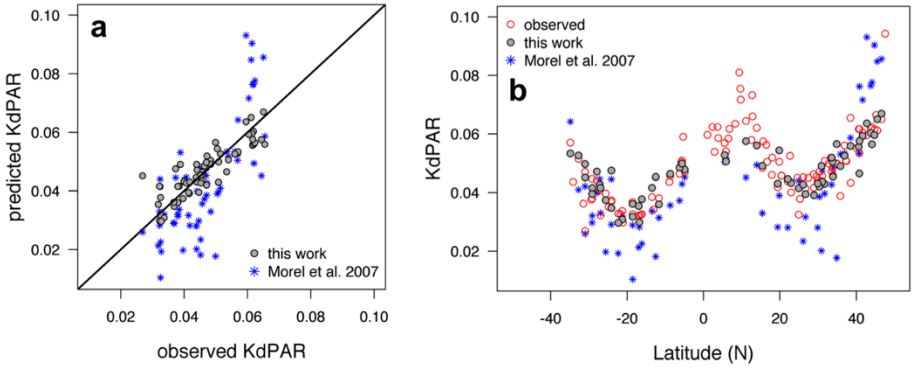

Figure S3. a) Comparison between observed attenuation coefficient for photosynthetically available radiation (K_dPAR') from CTD PAR profiles in AMTs 20-24 (red empty circles, locations displayed in **Fig. 1**) *versus* predicted K_dPAR using equation (1) of the present work (grey-filled circles) and using equation (9') from Morel et al. 2007 (blue stars); b) Observed (AMTs 20 to 24, n=108 observations) and predicted K_dPAR (n=61 observations) across the Atlantic Ocean. The number of predicted values is lower than those observed because of the lack of availability of $R_{rs}(443)$ information near the equatorial convergence caused by cloud cover.

To calculate the day length (*DL*), equation (S3) was taken from Forsythe et al. (1995):

$$DL = \left[\left(-1 \cos \frac{Jday}{57.295} \left(\sin \frac{Lat}{57.295} \right)^3 \right) + 1 \right] 12 \quad , \tag{S3}$$

where *Jday* is the Julian day (day of the year) and *Lat* is the latitude (degrees N). The absolute solar zenith angle at noon θ_s was derived from the sun declination (Forsythe et al. 1995) using equations (S4) and (S5):

$$\delta' = \left[23.45 \sin \left(\frac{360}{365} (284 + Jday) \right) \right] \left[\frac{\pi}{180} \right], \tag{S4}$$

$$\theta_s = Lat - \delta' \tag{S5}$$

where δ' is the sun declination (degrees), *Jday* is the day of the year, and *Lat* is the latitude (degrees). Finally, the depth z (in metres) associated with a particular *fPAR* was calculated using equation (S6) from Kirk (2011):

$$z = -\frac{\ln (fPAR(z))}{K_dPAR} \tag{S6}$$

4. Choice of a two-component model

In vertical profiles of *Prochlorococcus* cell abundance, the depth of the deep *Prochlorococcus* maximum (Z_{DPM}) is often located a few metres above the depth of the deep chlorophyll maximum (Z_{DCM}), particularly in subtropical gyre waters (**Fig. S7**). In our model, this deep *Prochlorococcus* maximum (*DPM*) is generated not from a peak in abundance of *ProI* or *ProII*, but from the sum of these two components. Therefore, Z_{DPM} is typically located at the depth where the *ProI* and *ProII* co-occur. When the *DCM* is shallower than 60 m, the maximum *Prochlorococcus* abundance in a profile often

occurs at the Z_{DCM} or at the sub-surface chlorophyll maximum because *ProII* is nearly absent (**Fig. S7a-c**).

To test if the deep component *ProII* is relevant to determine the vertical distribution of *Prochlorococcus* cells, curves of observed *Prochlorococcus* cell abundance versus *fPAR* were fitted to equation (S7) in each sampling station:

$$Pro_{total}(z) = Pro_{surf}{}' \left(1 - \exp\left(-\left(\frac{\alpha\ fPAR(z)}{Pro_{surf}{}'}\right)\right)\right) \quad \text{(S7)}$$

where Pro_{surf}' is the abundance of *Prochlorococcus* cells at the sea surface observed *in situ*, and is considered to be the maximum cell abundance *Prochlorococcus* can reach in a profile. In this equation, the deep maximum in the abundance of *Prochlorococcus* cells is neglected (*ProII* is absent).

To estimate the total *Prochlorococcus* abundance at each depth ($Pro_{total}(z)$) and the abundance of *Prochlorococcus* cells integrated at the surface 200 m of the water column (Pro_{int}) using satellite observables, *fPAR(z)* was calculated using equation (2), and Pro_{surf} was calculated using equations (3-5). The parameter α was taken from the fitted curves of *in-situ* observations using equation (S7), then averaged.

The comparison between values of Pro_{int} calculated using the two-component model ($Pro_{total}(z)$ = *ProI(z) + ProII(z)*) and the one-component model (where $Pro_{total}(z)$ is calculated using equation (S7)) is shown in **Fig. S4**. The one-component model ($r^2 = 0.36$) underestimates Pro_{int} in the gyres, especially the South Atlantic gyre, as these are the areas where the deep *Prochlorococcus* maximum is most important and dictates the patterns of integrated cell abundances (**Fig. S4a-b**). In these areas, predictions made by the two-component model ($r^2 = 0.48$) are more accurate (**Fig. S4c-d**). For this reason, we opted for the two-component model.

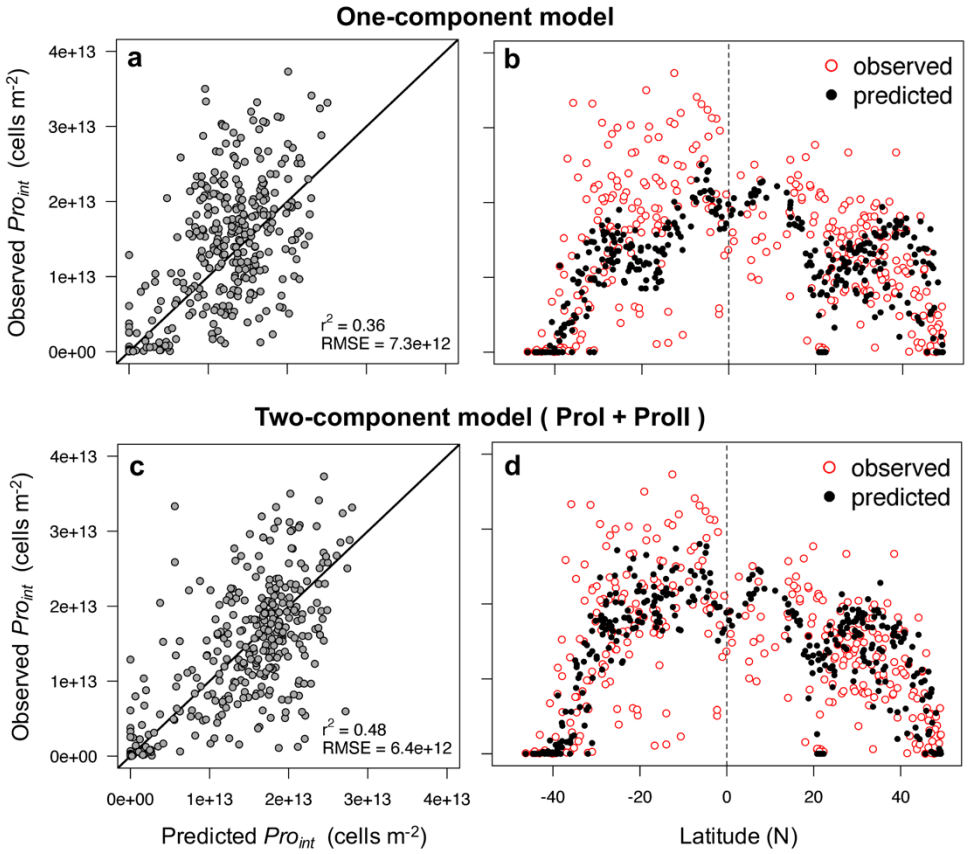

Figure S4. a) Comparison between observed and predicted *Prochlorococcus* cell abundance integrated in the water column (Pro_{int}) using the one-component model (equations (2), (3-5), and (S7)); b) Observed and predicted Pro_{int} across the Atlantic Ocean (locations displayed in **Fig. 1**) using the one-component model (AMT 12 to 24, n = 340 observations); c) Comparison between observed and predicted Pro_{int} using the two-component model (equations (1) to (15)); d) Observed and predicted Pro_{int} across the Atlantic Ocean (AMT 12 to 24, n = 340 observations) using the two-component model.

5. Global estimates of the abundance of Prochlorococcus cells

The validation of the two-component model in areas outside the Atlantic Ocean was performed using the PANGAEA dataset (Buitenhuis et al., 2012). Considering the Pacific and Indian Oceans, and the limited number of observations matching satellite data, the performance of the model to estimate the abundance of *Prochlorococcus* cells integrated in the water column (Pro_{int}^{3}) was poor when using exclusively satellite data ($r^2 = 0.14$) **(Table S1)**. However, when *in-situ* measurements of the abundance of *Prochlorococcus* at the sea surface were used to calculate the vertical distribution of cells using this same model, the performance of the model improved drastically ($r^2 = 0.75$ to Pro_{int} and 0.83 to Pro_{total}) **(Table S1)**.

Therefore, in areas outside the Atlantic Ocean, the model is compromised by the equation to estimate the abundance of cells at the sea surface (Pro_{surf}), which is empirical and tuned to Atlantic Ocean waters (as it was derived from the AMT dataset). This means it is possible to improve the performance of the model in other ocean basins by tuning the equation to calculate Pro_{surf} to these areas, which requires a

robust dataset containing vertical profiles of the abundance of *Prochlorococcus* cells (for validation), and the temperature at the sea surface and at the depth of 200 metres (used as input to the equation).

Table S1. Response variables with their respective centre-root (unbiased) mean square error (Δ), determination coefficient (r^2), and the number of observations (n) when compared with *in-situ* data from the PANGAEA dataset (3808 observations from 376 stations) (Buitenhuis et al., 2012). Variable definitions are listed in **Table 1**.

Variable	Equation	Δ PANGAEA	r^2 PANGAEA	n
$Pro_{total}(z)$ [2]	(9)	2.800×10^4	0.83	3808
Pro_{int} [2]	(10)	3.914×10^{12}	0.75	376
Pro_{surf} [3]	(3-5)	6.431×10^4	0.13	376
$Pro_{total}(z)$ [3]	(9)	6.471×10^4	0.38	3808
Pro_{int} [3]	(10)	7.601×10^{12}	0.14	376

Prochlorococcus cell abundances calculated using:
[2] *in-situ* Pro_{surf}' and modelled Z_{DCM} ;
[3] modelled Pro_{surf} and modelled Z_{DCM}.

The two-component model estimates a global *Prochlorococcus* standing stock of approximately 3.4×10^{27} cells representing a biomass of 171 Mt of carbon **(Table 3)**, which is equivalent to 65% of the global phytoplankton carbon standing stock in the mixed layer (Behrenfeld et al., 2015; Kostadinov et al., 2016; Roy et al., 2017), and higher than the carbon contained in all picophytoplankton in the mixed layer (140 Mt C), as reported by Roy et al. (2017), highlighting the importance of picophytoplankton below the mixed layer in oligotrophic waters. Most cells are accumulated within the subtropical gyres and at the Equatorial Convergence Zone **(Fig. S5a-c)**, and reside within the top 200 m of the surface ocean **(Fig. S5; Fig. S7)**, with 43% of the cells accumulated on the top 45 metres of the water column, and 57% between the depths of 45 and 200 m.

A marked seasonal cycle in the global cell stock was observed from 2003 to 2014 (**Fig. S6a**), but no significant long-term temporal trend was evident **(Figs. S6b)**. At the ocean's surface, where the global seasonal signal is strong **(Fig. S6c)**, a slight decrease in cell abundance was observed from 2003 to 2011, followed by an increase from 2012 to 2014 **(Fig. S6d)**. These estimates of global *Prochlorococcus* cell and carbon stocks are 5% lower than that made by Williams and Follows (2011) (3.5×10^{27} cells, 180 Mt C), and 22% higher than the estimate of Flombaum et al. (2013) (2.9×10^{27} cells).

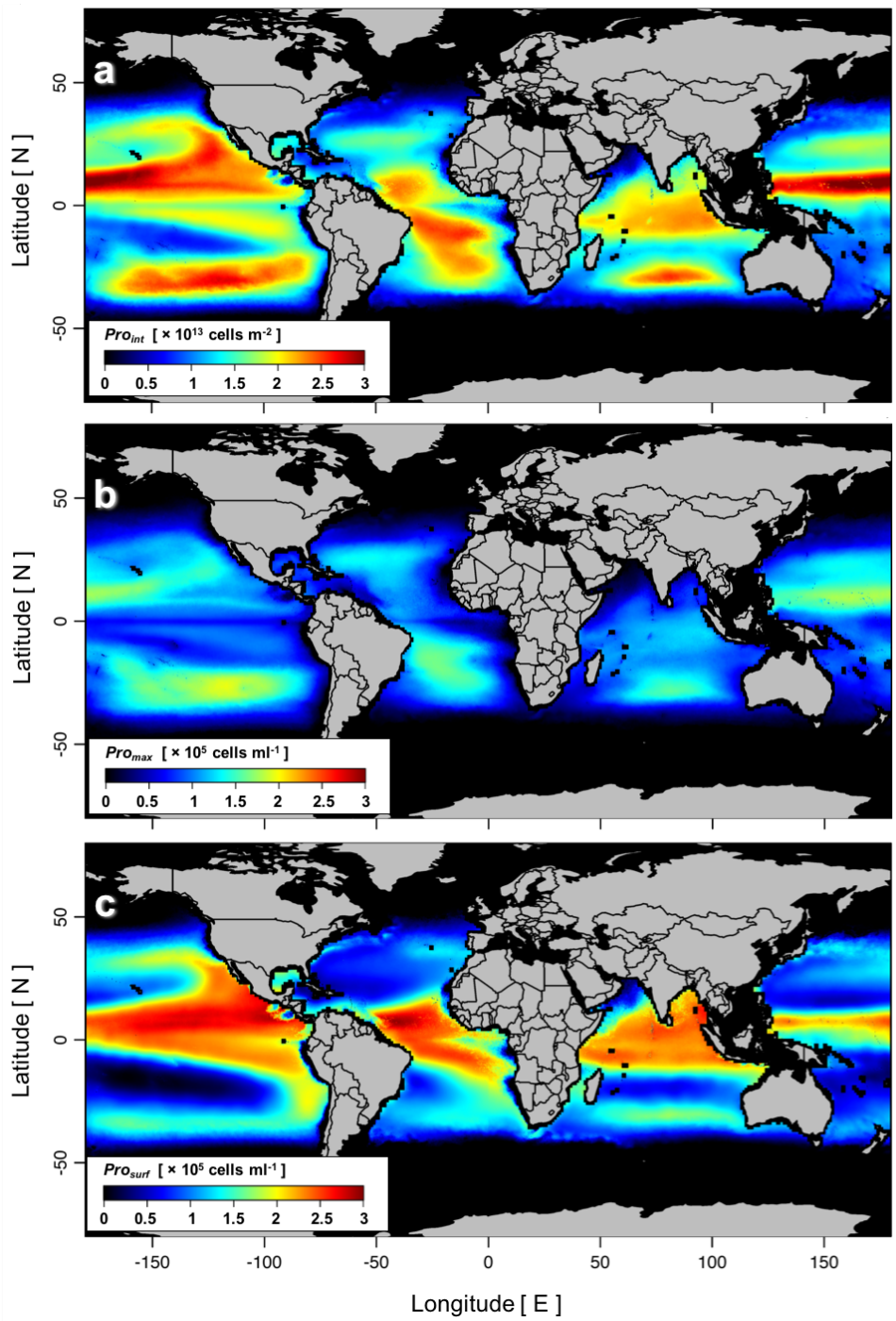

Figure S5. Annual mean (averaged from 2003 to 2014) of a) the predicted *Prochlorococcus* cell abundance integrated in the top 200 m of the water column (Pro_{int}^{3}) using the complete model (equations (1) to (10) of the present work); b) predicted maximum *Prochlorococcus* cell abundance at depth (Pro_{max}^{3}) using equation (7); and c) predicted *Prochlorococcus* cell abundance at surface (Pro_{surf}^{3}) from equations (3) to (5).

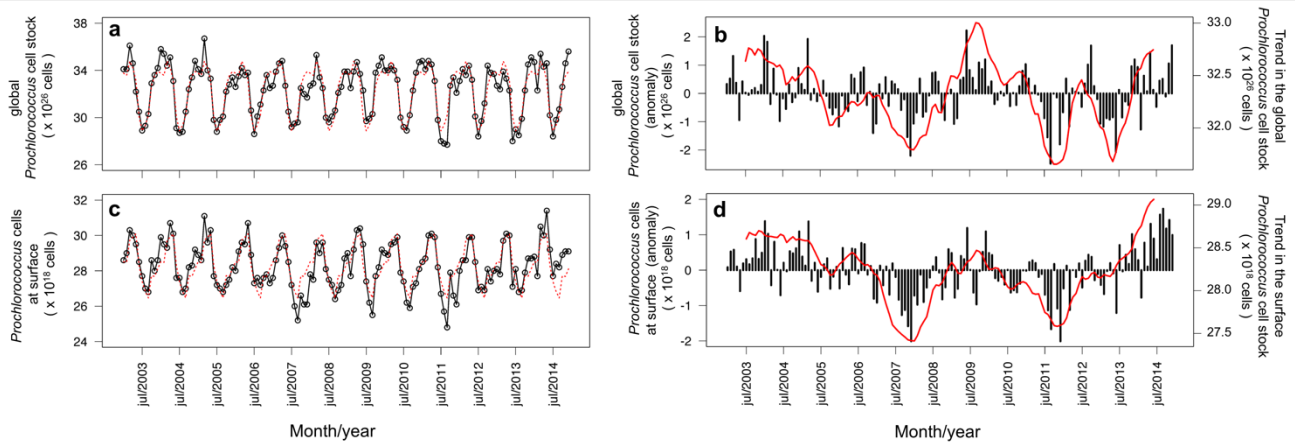

Figure S6. Time series of a) global horizontally- and vertically-integrated *Prochlorococcus* cell stocks and c) horizontally-integrated cell abundance near the ocean surface, with corresponding anomalies and trends (b,d). Values were computed using monthly-composite input variables from the year 2002 to 2014 (NASA Ocean Color 2014). Anomalies were calculated by subtracting the monthly climatology (dashed red lines in figures (a) and (c)) from the calculated values for each month (black lines in figs. (a) and (c)). Red lines in figures (b) and (d) show the moving trend.

Additional Tables

Table S1. AMT Cruises used in this study and corresponding dates.

cruise	dates		
AMT 12	12.05.2003	-	17.06.2003
AMT 13	08.09.2003	-	13.10.2003
AMT 14	26.04.2004	-	2.06.2004
AMT 15	19.09.2004	-	29.10.2004
AMT 16	19.05.2005	-	29.06.2005
AMT 17	15.10.2005	-	28.11.2005
AMT 18	3.10.2008	-	10.11.2008
AMT 19	13.10.2009	-	1.12.2009
AMT 20	12.10.2010	-	25.11.2010
AMT 21	29.09.2011	-	14.11.2011
AMT 22	10.10.2012	-	24.11.2012
AMT 23	7.10.2013	-	8.11.2013
AMT 24	21.09.2014	-	6.11.2014

Table S2. AMT cruises used to generate models to calculate the light attenuation coefficient K_dPAR, depth of the deep chlorophyll maximum Z_{DCM}, *Prochlorococcus* cell abundance at surface Pro_{surf}, and at the deep maximum Pro_{max}. n = number of observations (for parameterization and validation).

variable	AMT cruises for parameterization	n param.	AMT cruises for validation	n valid.
K_dPAR	20 , 21 , 22 , 23 , 24	3421	12 – 20	5842
Z_{DCM}	12 , 13 , 17 , 18 , 20 , 22 , 24	4575	14, 15, 16, 19, 21, 23	4147
Pro_{surf}	12 , 13 , 17 , 18 , 20 , 22 , 24	4575	14, 15, 16, 19, 21, 23	4147
Pro_{max}	12 , 13 , 17 , 18 , 20 , 22 , 24	4575	14, 15, 16, 19, 21, 23	4147

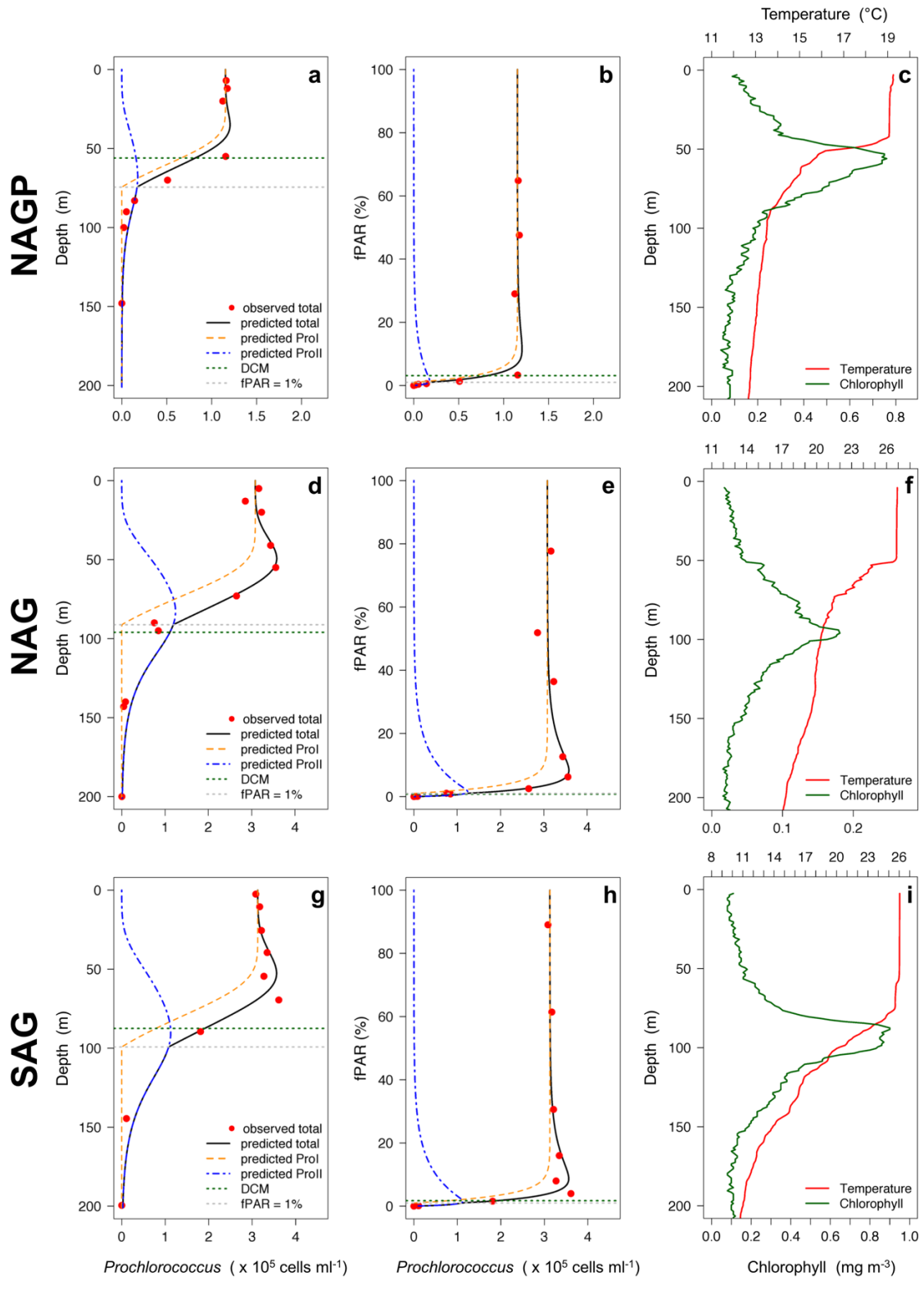

Figure S7. Vertical profiles of observed and estimated *Prochlorococcus* cell abundances (a,d,g) over depth and (b,e,h) over the fractional PAR *fPAR*, with (c,f,i) corresponding profiles of temperature and chlorophyll from CTD casts at a site in (a-c) the North Atlantic Gyre periphery (NAGP), (d-f) inside the North Atlantic Gyre, and (g-i) the South Atlantic Gyre. *In-situ* observations are represented by red dots, predicted profiles of *ProI* by the orange dashed line, predicted profiles of *ProII* by the blue dashed line, and predicted profiles of total *Prochlorococcus* abundance by the solid black line. Data from AMT 24 (2014).

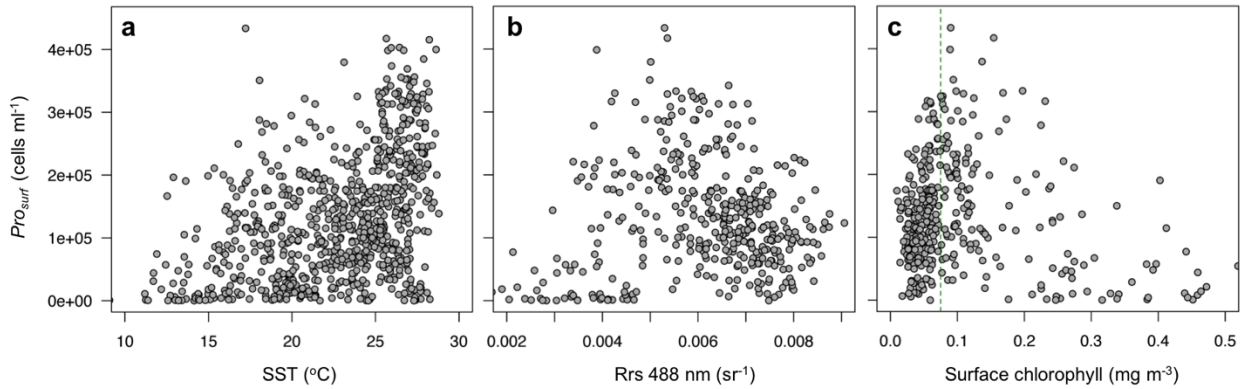

Figure S8. Relationship between *Prochlorococcus* cell abundance at the sea surface (Pro_{surf}) and: a) sea surface temperature (*SST*), b) remote sensing reflectance at 448 nm ($R_{rs}(488)$) and c) surface chlorophyll concentration on AMTs 12 to 24. The dashed line indicates the chlorophyll threshold used to delineate the subtropical gyres.

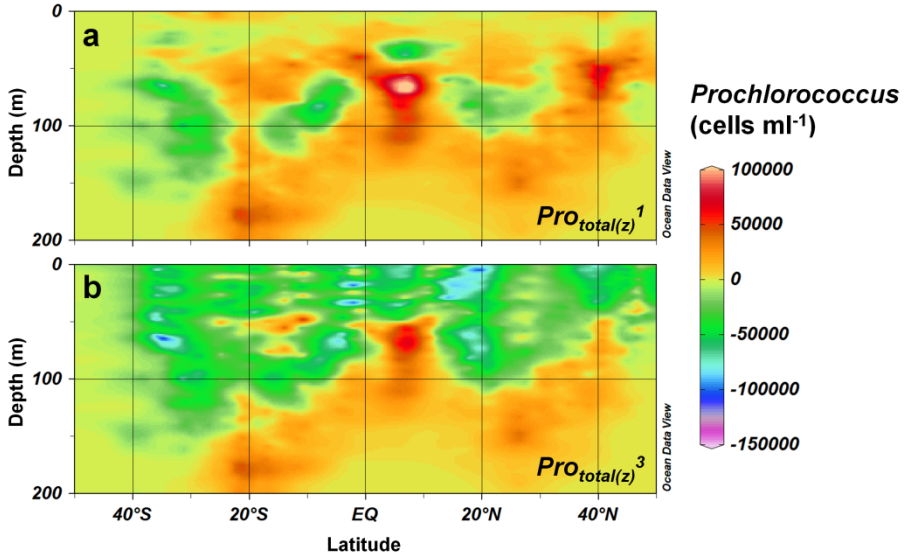

Figure S9. Difference (residue) between predicted and observed *Prochlorococcus* cell abundance over depth ($Pro_{total(z)}$) a) using the partial model where *in-situ* observations of Z_{DCM}, K_dPAR, and Pro_{surf} are used as input to calculate $Pro_{total(z)}^{1}$ (i.e. equations (1), (3-5), and (7) are excluded); and b) using the full model with remote-sensing input ($Pro_{total(z)}^{3}$ from equations (1) to (10)).

References

Behrenfeld, M.J., O'Malley, R.T., Boss, E.S., Westberry, T.K., Graff, J.R., Halsey, K.H., Milligan, A.J., Siegel, D.A., Brown, M.B., 2015. Revaluating ocean warming impacts on global phytoplankton. Nat. Clim. Chang. 6, 323–330. doi:10.1038/nclimate2838

Buitenhuis, E.T., Li, W.K.W., Vaulot, D., Lomas, M.W., Landry, M.R., Partensky, F., Karl, D.M., Ulloa, O., Campbell, L., Jacquet, S., Lantoine, F., Chavez, F., Macias, D., Gosselin, M., McManus, G.B., 2012. Picophytoplankton biomass distribution in the global ocean. Earth Syst. Sci. Data 4, 37–46. doi:10.5194/essd-4-37-2012

Hijmans, R.J., 2016. raster: Geographic data analysis and modelling. R package version 2.5-8.

James, D., Hornik, K., 2017. chron: Chronological objects which can handle dates and times. R package version 2.3-50.

Kirk, J.T.O., 2011. Light and Photosynthesis in Aquatic Ecosystems, 3rd ed. Cambridge University Press, Cambridge.

Kostadinov, T.S., Milutinovi, S., Marinov, I., Cabr??, A., 2016. Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution. Ocean Sci. 12, 561–575. doi:10.5194/os-12-561-2016

McClain, C.R., Signorini, S.R., Christian, J.R., 2004. Subtropical gyre variability observed by ocean-color satellites. Deep. Res. Part II Top. Stud. Oceanogr. 51, 281–301. doi:10.1016/j.dsr2.2003.08.002

Morel, A., Huot, Y., Gentili, B., Werdell, P.J., Hooker, S.B., Franz, B.A., 2007. Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. Remote Sens. Environ. 111, 69–88. doi:10.1016/j.rse.2007.03.012

NASA, 2014. Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua) Ocean Color Data. Goddard Space Flight Center Ocean Biology Processing Group, USA.

Nychka, D., Furrer, R., Paige, J., Sain, D., 2015. fields: Tools for spatial data. R package version 8.10. doi:10.5065/D6W957CT

Olson, R.J., Zettler, E.R., DuRand, M.D., 1993. Phytoplankton analysis using flow cytometry, in: Kemp, P.F., Sherr, E.B., Cole, J.J. (Eds.), Handbook of Methods in Aquatic Microbial Ecology. Lewis Publishers, Boca Raton, USA, pp. 175–186.

Pierce, D., 2015. ncdf4: Interface to unidata netCDF (Version 4 or Earlier) format data files. R package version 1.15.

R Core Team, 2017. R: A language and environment for statistical computing. Viena.

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

Sarkar, D., 2008. Lattice: Multivariate data visualization with R. Springer, New York.

Wickham, H., 2009. ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York.

Wickham, H., Chang, W., 2016. devtools: Tools to make developing R packages easier. R package version 1.12.0.

Zubkov, M. V., Burkill, P.H., Topping, J.N., 2007. Flow cytometric enumeration of DNA-stained oceanic planktonic protists. J. Plankton Res. 29, 79–86. doi:10.1093/plankt/fbl059

Zubkov, M. V., Sleigh, M.A., Burkill, P.H., Leakey, R.J.G., 2000. Picoplankton community structure on the Atlantic Meridional Transect: a comparison between seasons. Prog. Oceanogr. 45, 369–386. doi:10.1016/S0079-6611(00)00008-2

Zubkov, M. V, Burkill, P.H., 2006. Syringe Pumped High Speed Flow Cytometry of Oceanic Phytoplankton 1019, 1010–1019. doi:10.1002/cyto.a