

2019

# A Satellite Assessment of Environmental Controls of Phytoplankton Community Size Structure

Colleen B. Mouw

*University of Rhode Island*, [cmouw@uri.edu](mailto:cmouw@uri.edu)

Audrey B. Ciochetto

*University of Rhode Island*

*See next page for additional authors*

Follow this and additional works at: <https://digitalcommons.uri.edu/gsofacpubs>

**The University of Rhode Island Faculty have made this article openly available.  
Please let us know how Open Access to this research benefits you.**

This is a pre-publication author manuscript of the final, published article.

Terms of Use

This article is made available under the terms and conditions applicable towards Open Access Policy Articles, as set forth in our [Terms of Use](#).

## Citation/Publisher Attribution

Mouw, C. B., Ciochetto, A. B., & Yoder, J. A. (2019). A satellite assessment of environmental controls of phytoplankton community size structure. *Global Biogeochemical Cycles*, 33. <https://doi.org/10.1029/2018GB006118> Available at: <https://doi.org/10.1029/2018GB006118>

This Article is brought to you for free and open access by the Graduate School of Oceanography at DigitalCommons@URI. It has been accepted for inclusion in Graduate School of Oceanography Faculty Publications by an authorized administrator of DigitalCommons@URI. For more information, please contact [digitalcommons@etal.uri.edu](mailto:digitalcommons@etal.uri.edu).

---

**Authors**

Colleen B. Mouw, Audrey B. Ciochetto, and James A. Yoder

1                                   **A Satellite Assessment of Environmental Controls**  
2                                   **of Phytoplankton Community Size Structure**  
3

4                                   **Colleen B. Mouw<sup>1</sup>, Audrey B. Ciochetto<sup>1</sup>, and James A. Yoder<sup>1</sup>**

5                                   <sup>1</sup>University of Rhode Island, Graduate School of Oceanography  
6                                   215 South Ferry Road, Narragansett, RI, 02882, USA  
7

8                                   Corresponding author: Colleen Mouw ([cmouw@uri.edu](mailto:cmouw@uri.edu)), +1-401-874-6506  
9

10   **Key Points:**

- 11                   • Globally, light availability in the water column is the most important parameter for  
12                   phytoplankton size distribution
- 13                   • Regionally, phytoplankton size distributions vary, responding to variable light and modes  
14                   of nutrient delivery
- 15                   • Cell size is increasing in the cold ocean and the dynamic regions in the warm ocean and  
16                   declining in the warm ocean  
17  
18  
19  
20  
21

22 **Abstract**

23 Phytoplankton play a key role as the base of the marine food web and a crucial component in the  
24 Earth's carbon cycle. There have been a few regional studies that have utilized satellite-estimated  
25 phytoplankton functional type products in conjunction with other environmental metrics. Here we  
26 expand to a global perspective and ask, what are the physical drivers of phytoplankton composition  
27 variability? Using a variety of satellite observed ocean color products and physical properties  
28 spanning 1997-2015, we characterize spatial and temporal variability in phytoplankton community  
29 size structure in relation to satellite-based physical drivers. We consider the relationships globally  
30 and by major thermal regimes (cold and warm), dominant size distribution, and chlorophyll  
31 concentration variability. Globally, euphotic depth is the most important parameter driving  
32 phytoplankton size variability and also over the majority of the high latitude ocean and the central  
33 gyres. In all other regions, size variability is driven by a balance of light and mode of nutrient  
34 delivery. We investigated the relationship between size composition and chlorophyll  
35 concentration and the physical drivers through correlation analysis. Changes in size composition  
36 over time are regionally varying and explained by temporal shifts in the varying physical  
37 conditions. These changes in phytoplankton size composition and the varying underlying  
38 physical drivers will ultimately impact carbon export and food web processes in our changing  
39 ocean.

40

41

## 42 **1 Introduction**

43 Phytoplankton play a key role as the base of the marine food web and are a crucial  
44 component in the earth's carbon cycle. Given this importance, there have been many studies that  
45 have characterized phytoplankton distributions from field observations (Boyd et al., 2010; Brun et  
46 al., 2015), satellite estimates (McClain et al., 2009; Siegel et al., 2013; Mouw et al., 2017) and  
47 modeling (Dutkiewicz et al., 2013; Henson et al., 2017). Previous investigations have suggested  
48 alterations in biomass, productivity, and community composition as a result of the changing  
49 environment (Bopp et al., 2005; Behrenfeld et al., 2006; Vantrepotte and Melin, 2009; Siegel et  
50 al., 2013; Gregg & Rousseaux, 2014; Rousseaux & Gregg, 2015; Dutkiewicz et al., 2013). These  
51 changes are regionally variable and complex with multiple drivers at play at once, yet signatures  
52 of trends are being detected over the timeframes that satellites have observed (e.g. Henson et al.,  
53 2017).

54 Considering more than just abundance, phytoplankton community composition plays an  
55 important role in the intricacies of food web dynamics and their influence on carbon export flux  
56 (Guidi et al., 2015; Mouw et al., 2016). It is well understood that small cells are more commonly  
57 dominant in stratified, high light conditions, while large cells dominate in well mixed, lower light  
58 conditions (Chisholm, 1992). Now with a variety of approaches to estimate phytoplankton  
59 community structure from satellite (Mouw et al., 2017; IOCCG, 2014), we have the ability to take  
60 a broader look beyond abundance. A variety of approaches have emerged that attempt to  
61 discriminate phytoplankton functional types (PFT), which include algorithms that retrieve  
62 phytoplankton size classes (PSC), phytoplankton taxonomic composition (PTC), or particle size  
63 distribution (PSD). A PFT is an aggregation of phytoplankton, where irrespective of their  
64 phylogeny, they share similar biogeochemical or ecological roles. The existing approaches vary  
65 in what phytoplankton groupings they retrieve and the underlying mechanisms in which they  
66 derive group information (Mouw et al., 2017).

67 There have been a few studies that have utilized satellite PFT products in conjunction with  
68 other environmental metrics. Thus far, these have been regional in scope. In high latitudes, an  
69 increase in diatoms were observed during positive phases of local climate indices, suggesting the  
70 increases were driven by nutrient supply (Alvian et al., 2013). Coccolithophore abundance was  
71 found to correspond to shallow mixed layer with, low wind speed, and increasing sea surface  
72 temperature (Sadeghi et al., 2012). In the Indian Ocean, the variance in phytoplankton structure  
73 was explained by sea surface height, stratification and sea surface temperature (Brewin et al.,  
74 2012). Southern Ocean diatom phenology was found to be driven by the polar front, ice extent  
75 and oppositely correlated with El Niño Southern Oscillation and the Southern Annual Mode  
76 (Soppa et al., 2016).

77 Here we expand to a global view to assess the physical drivers of phytoplankton  
78 composition variability from satellite products. How have phytoplankton (i.e. chlorophyll  
79 concentration and composition) distributions changed over the satellite record? What are the  
80 physical drivers of this variability? Over the satellite record, we characterize the relationship  
81 between chlorophyll *a* concentration and phytoplankton composition and the variability of  
82 phytoplankton distributions to define regions based on persistent patterns. We then determine the  
83 dominant physical processes that are responsible for the observed variability and change over time.

## 84 2 Materials and Methods

### 85 2.1 Imagery and Reanalysis Products

86 A summary of data products, descriptions and sources, including website links, can be found in  
87 Table 1.

#### 88 2.1.1 Ocean Color Imagery

89 To allow the greatest length of time of continuous ocean color imagery, we utilized merged  
90 imagery products obtained from the Ocean Color Climate Change Initiative (OC-CCI, version 3.0,  
91 the latest version at the time of analysis, Grant et al., 2017). OC-CCI has globally merged Sea-  
92 Viewing Wide Field-of-View Sensor (SeaWiFS), Medium Resolution Imaging Spectrometer  
93 (MERIS), Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua), and Visible Infrared  
94 Imaging Radiometer Suite (VIIRS) imagery for a continuous record from 1997 through 2015  
95 (Sathyendranath et al., 2018). We utilized monthly resolution, 4 km products. The OC-CCI  
96 products used here include: chlorophyll *a* concentration ([Chl],  $\text{mg m}^{-3}$ ), spectral remote sensing  
97 reflectance ( $R_{rs}(\lambda)$ ,  $\text{sr}^{-1}$ ), spectral dissolved and detrital absorption ( $a_{dg}(\lambda)$ ,  $\text{m}^{-1}$ ), and the diffuse  
98 attenuation coefficient at 490 nm ( $K_d(490)$ ,  $\text{m}^{-1}$ ). OC-CCI derives  $K_d(490)$  from the Lee et al.  
99 (2005) algorithm, which is independent of [Chl]. Here,  $K_d(490)$  was used to calculate euphotic  
100 depth ( $z_{eu} = 4.6/K_d(490)$ , Morel & Berthon 1989). OC-CCI provides the spectral products at  
101 the SeaWiFS bands, by band-shifting the  $R_{rs}(\lambda)$  values from MERIS, MODIS and VIIRS to match  
102 those of SeaWiFS. The OC-CCI [Chl] product is estimated by a blended combination of the  
103 empirical band ratio algorithm (OCx, O'Reilly et al., 1998), and the ocean color index algorithm  
104 (OCI) which itself blends the band ratio algorithm and color index (CI, Hu et al., 2012) (Jackson  
105 & Grant, 2016). The OC-CCI  $a_{dg}(\lambda)$  product is calculated from the quasi semi-analytical algorithm  
106 (Lee et al., 2002; 2007).

107 Satellite PFT algorithms have a variety of phytoplankton product outputs and units (Mouw  
108 et al., 2017; IOCCG, 2014). This presents an additional layer of challenge, precluding direct  
109 comparison of algorithm performance. Instead, metrics such as phenological cycle, have been  
110 used as a way to inter-compare PFT algorithms (Kostadinov et al., 2017). This intercomparison  
111 revealed that while PFT algorithms agree across broad scales, they do not all agree under all  
112 circumstances. Here we sought to utilize a PFT product that performed near the mean of the  
113 phenological metrics (phenological shape, magnitude and month of maximum) that Kostadinov et  
114 al. (2017) assessed, as well as that with high validation metrics reported from the original  
115 publication (compiled by Mouw et al., 2017). Further, phytoplankton size is one of the best  
116 characterized traits structuring food webs due to many ecosystem and physiological processes that  
117 are mediated by size such as: nutrient acquisition and utilization, light acquisition, sinking, and  
118 grazer interactions (Finkel, 2007; Litchman & Klausmeier, 2008; Finkel et al., 2009; Litchman et  
119 al., 2010; Wirtz, 2012). We selected the satellite output from Mouw and Yoder (2010) that  
120 estimates phytoplankton size class as percent microplankton ( $S_{fm}$ ,  $> 20 \mu\text{m}$ ). The uncertainty  
121 metrics of this product are one of the best performing of the PFT algorithms reviewed by Mouw  
122 et al. (2017) with  $r^2=0.6$  and RMSE of 12.64. The calculation of  $S_{fm}$  requires satellite  $R_{rs}(\lambda)$ , [Chl],  
123 and  $a_{dg}(\lambda)$ . These are taken from the OC-CCI products described above. This is an absorption-  
124 based approach where the chlorophyll-specific absorption spectra for phytoplankton size class  
125 extremes, pico- ( $0.2\text{--}2 \mu\text{m}$ ) and microplankton ( $> 20 \mu\text{m}$ ), are weighted by  $S_{fm}$  (Ciotti et al., 2002;  
126 Ciotti & Bricaud, 2006).  $S_{fm}$  is estimated from a look-up table containing simulated [Chl],  
127  $a_{dg}(443)$ ,  $R_{rs}(\lambda)$ , and  $S_{fm}$ . For a given pixel, satellite-estimated [Chl] and  $a_{dg}(443)$ , are used to  
128 narrow the search space within the look-up table. Of the remaining options, the closest simulated

129  $R_{rs}(\lambda)$  to the satellite-observed  $R_{rs}(\lambda)$  is selected and the associated  $S_{fm}$  is assigned. The  $S_{fm}$  product  
 130 processed from OC-CCI imagery is available on PANGAEA:  
 131 <https://doi.pangaea.de/10.1594/PANGAEA.892211>.

132 We utilized the vertically generalized primary production model (VGPM) as the net  
 133 primary productivity (NPP) product (Behrenfeld & Falkowski, 1997). NPP is a function of  
 134 chlorophyll, available light, and temperature-dependent photosynthetic efficiency. We accessed  
 135 the monthly, 9 km VGPM NPP imagery for SeaWiFS and MODIS (derived from R2014  
 136 processing). Following the recommendations of Mélin (2016) to prevent the introduction of long-  
 137 term anomalous trends from cross-mission differences, SeaWiFS and MODIS data were merged  
 138 from a bias-corrected signal. Briefly, monthly climatologies for each pixel were created from  
 139 years with mission overlap (2003-2007) for SeaWiFS and MODIS records. Gaps of less than two  
 140 months in the climatology were filled using a spline interpolation. The entire MODIS record was  
 141 then adjusted by the difference between SeaWiFS and MODIS climatologies:

$$142 \quad x_{a,corr}(p, m) = x_a(p, m) + x_{s,clim}(p, m) - x_{a,clim}(p, m)$$

143 where  $x_{a,corr}$  is the corrected MODIS signal for a given pixel and month,  $x_a$  is the original MODIS  
 144 signal, and  $x_{s,clim}$  and  $x_{a,clim}$  are the SeaWiFS and MODIS climatologies, respectively. The final  
 145 combined record was created by averaging SeaWiFS with the bias-corrected MODIS signal. The  
 146 pre-MODIS time period includes data from SeaWiFS alone and the post-SeaWiFS time period  
 147 comprises  $x_{a,corr}$  alone.

148

#### 149 2.1.2 Physical Data Sets:

150 Several physical products were utilized to diagnose the drivers of phytoplankton  
 151 community variability. Satellite and blended products were used to characterize  
 152 photosynthetically active radiation (PAR,  $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ ), sea surface temperature (SST,  $^{\circ}\text{C}$ )  
 153 and sea-level anomaly (SLA, m). PAR is the quantum energy flux from the sun between 400 and  
 154 700 nm and is a standard product hosted on the NASA Ocean Color web  
 155 (<https://oceancolor.gsfc.nasa.gov/>) at monthly, 9 km resolution. PAR from SeaWiFS and MODIS  
 156 missions (R2014 processing) were merged following Mélin (2016) as described above. The Group  
 157 for High Resolution Sea Surface Temperature (GHRSSST) retrieves SST products that are hosted  
 158 by the National Oceanographic Data Center (NODC). We utilized the Level 4 global product,  
 159 which provides gap-free data at quarter-degree, daily resolution by combining *in situ* sensors with  
 160 satellite products from the Advanced Very High-Resolution Radiometer (AVHRR) Pathfinder  
 161 missions (Version 5 processing) ([www.ghrsst.org](http://www.ghrsst.org)). SLA represents the difference in sea-level  
 162 height from a reference period. The SSALTO/Data Unification and Altimeter Combination  
 163 System (DUACS) hosted by Archiving, Validation and Interpretation of Satellite Oceanographic  
 164 data (AVISO) is a multi-sensor satellite product derived from seven international satellite missions  
 165 (Sarat/AltiKa, Jason-1 and -2, Cryosat-2, Envisat, ERS-1 and -2, GFO and HY-2A). SLA is  
 166 retrieved in quarter-degree, monthly resolution relative to the twenty-year mean profile from 1993-  
 167 2012. The seasonal cycle is not removed from the dataset.

168 Reanalysis data products were used to retrieve net total heat flux ( $Q_{net}$ ,  $\text{W m}^{-2}$ ), mixed layer  
 169 depth (MLD, m) and stratification index ( $\Delta\rho_{200}$ ,  $\text{kg m}^{-3}$ ). Net total heat flux is used to determine  
 170 if the ocean is a source or sink for heat energy. The National Centers for Environmental Prediction  
 171 (NCEP) and the National Center for Atmospheric Research (NCAR) provide reanalysis products  
 172 generated from a variety of satellite, airborne and *in situ* platforms, for latent and sensible heat  
 173 fluxes on a T62 Gaussian grid with monthly resolution (Kalnay et al., 1996). Surface fluxes for  
 174 net latent heat flux, net longwave radiation, net shortwave radiation and sensible heat flux were

175 summed to retrieve  $Q_{\text{net}}$ . We utilized MLD and potential density ( $\sigma_{\theta}$ ,  $\text{kg m}^{-3}$ ) from the Simple  
 176 Ocean Data Assimilation (SODA, version 3.3.1, Carton, et al., 2018), which is forced by the  
 177 Modern-Era Retrospective analysis for Research and Applications (MERRA-2) dataset. SODA  
 178 assimilates a variety of *in situ* and satellite observations with a model framework to reconstruct  
 179 the 3-D physical history of the ocean on a half-degree grid with 50 depth levels ranging from 5 m  
 180 to 5000 m. For the density-based MLD product, the mixed layer is defined as the depth where  
 181 density exceeds surface density by  $0.03 \text{ kg m}^{-3}$ . We retrieved the stratification index from  $\sigma_{\theta}$  as  
 182 the difference in density between the surface and 200 m (Behrenfeld et al., 2006; Brewin et al.,  
 183 2012). We also considered bathymetry and wind speed, but did not find compelling relationships,  
 184 thus they were left out of further analysis and not reported here.

185

## 186 2.2 Data Processing Procedures:

187 The original downloaded satellite and modeled data products described above have a  
 188 variety of gridding, spatial resolution and time scales. In order to directly relate one product to  
 189 another, we created a uniform  $1 \times 1$  degree, gap-free time series for each data product based on the  
 190 processing steps of Yoder and Kennelly (2003). First, derived products were retrieved from the  
 191 original dataset (Step 1). Next, data were spatially smoothed or re-gridded to a  $1 \times 1$  degree product  
 192 (Steps 2 and 3) followed by log transformation where appropriate (Step 4). Finally, data were  
 193 temporally smoothed and filled in an attempt to produce a gap-free time series (Steps 5 and 6) and  
 194 quality controlled (Step 7). Details of these processing steps are as follows:

- 195 1) Derived products ( $S_{\text{fm}}$ ,  $Z_{\text{eu}}$ , and  $\Delta\rho_{200}$ ) were retrieved, or SeaWiFS and MODIS signals  
 196 were combined for NPP and PAR products following Mélin (2016).
- 197 2) Data were spatially smoothed to  $1/4$  degree via geometric mean for 4 km and 9 km  
 198 products, or daily images for SST were combined to create a monthly mean product.
- 199 3) Data were spatially smoothed to 1 degree via median filter, or data were re-gridded to 1  
 200 degree for  $Q_{\text{net}}$  product via bi-linear interpolation, or data were re-gridded to 1 degree for  
 201 MLD and  $\Delta\rho_{200}$  products via geometric mean.
- 202 4) Products with non-normal distributions were base-10 log transformed ( $S_{\text{fm}}$ , [Chl], NPP,  
 203 MLD and  $\Delta\rho_{200}$ ). Normality was assessed by comparing the skewness of the original  
 204 dataset to log transformed values.
- 205 5) Data were temporally smoothed via a 3-month moving average.
- 206 6) Gaps of 6 months or less were filled via spline interpolation. Gaps ranging from 5 to 6  
 207 months existed in, at most, 5% of global pixels for any given variable and were  
 208 concentrated at the very northern and southern most edges of the dataset.
- 209 7) Outliers greater than 5 standard deviations from the mean were removed.

210 The final data are  $1^{\circ}$  by  $1^{\circ}$  latitude/longitude bins with monthly resolution from January 1998 to  
 211 March 2015. Only those pixels with 100% data coverage for all data products were used in further  
 212 analysis; this includes almost all pixels between  $50^{\circ}\text{N}$  and  $50^{\circ}\text{S}$ .

213

## 214 2.3 Analysis

215 Long-term trends and correlation are used to understand temporal and spatial variability of  
 216 the dataset. To determine the long-term trend, the monthly climatological cycle for each pixel is  
 217 first removed from the dataset. The remaining linear trend was calculated using the Theil-Sen  
 218 approach, which is a non-parametric method insensitive to outliers where slope is retrieved as the  
 219 median of the distribution of slopes between every pair of points in the data set (Barton, Lozier &

220 William, 2015). Bayes factors ( $BF_{10}$ , unitless) were calculated to assess fit significance (Wetzels  
221 & Wagenmakers, 2012). Bayes factors represent the likelihood that a slope should be included in  
222 the model (slope is non-zero) versus that it should not (slope is zero). For example,  $BF_{10}=10$   
223 means the retrieved slope is ten times more likely to exist than a slope of zero. Here, we only  
224 present results with a  $BF_{10}>3$ , which is considered the cutoff for “substantial evidence” that a slope  
225 exists (Wetzels & Wagenmakers, 2012). Correlation between  $S_{fm}$  and each of the parameters was  
226 determined with Kendall’s rank correlation coefficient. Prior to retrieving correlation, products  
227 were standardized by subtracting the mean and dividing by the standard deviation to express them  
228 on the same scale. Again, only significant correlations,  $BF_{10}>3$ , are reported, with  $BF_{10}$   
229 representing the likelihood that a correlation exists versus that it does not.

230 Partial least squares regression (PLSR) with 10-fold cross validation is used to determine  
231 the relative importance of each parameter to  $S_{fm}$  (Wold et al., 2001). Again, data are standardized  
232 prior to analysis to express them on the same scale. PLSR combines predictor variables into  
233 principle components that are then regressed with  $S_{fm}$ . The method allows co-linearity between  
234 predictors since they all contribute to forming principle components. VIP (variable influence on  
235 projection) scores represent the relative importance of each predictor to  $S_{fm}$  variability, while  
236 regression coefficients indicate the magnitude and direction of the relationship with  $S_{fm}$ . For a  
237 given predictor, the VIP score quantifies the cumulative contribution of that predictor to each  
238 principle component weighted by the proportion of variance in  $S_{fm}$  explained by that component  
239 (Mehmood et al., 2012). Here, we consider results with  $VIP>0.5$  to be significant (Wold et al.,  
240 2001). Since data were standardized, the relative magnitude of regression coefficients also reflects  
241 the importance of each predictor to  $S_{fm}$ .

242 With PLSR, there is the possibility of finding significant correlation by chance, although  
243 this likelihood decreases as the dataset gets larger (Clark & Cramer, 1993). We performed a  
244 bootstrap test with the global dataset, where the rows of each predictor variable (i.e.  
245 latitude/longitude locations and times) were randomly paired with  $S_{fm}$  estimates prior to  
246 performing a PLSR. We repeated this process 1000 times and none of these cases explained a  
247 significant portion of the variance in  $S_{fm}$  or had significant VIP scores for any of the randomly  
248 ordered predictors. Thus, we are confident that our results with the ordered dataset are more than  
249 chance.

250 To assess confidence in parameter importance, we applied leave-one-predictor-out  
251 validation (Martens & Martens, 2000). This method repeats the PLSR analysis  $n+1$  times, where  
252  $n$  is the number of predictor variables. The first run includes all predictor variables and subsequent  
253 runs remove a single predictor from the dataset each time. Results are presented as the mean  
254 coefficients and VIP scores from the resulting distribution with error bars representing minimum  
255 and maximum values in the leave-one-predictor-out analysis. This is more appropriate for our  
256 large dataset than a jack-knife leave-one-replicate-out approach, where each data point is  
257 successively removed from the repeated analysis, as single measurements are not likely to alter  
258 final relationships in large data sets (Wold et al., 2001).

259

## 260 **3 Results**

### 261 3.1 Global Analysis:

262 The great advantage of using satellite products and merging them over time is the ability  
263 to explore temporal and spatial variability and the interrelation of these trends. Which parameters

264 display the greatest change over the satellite record and which show significant correlation with  
 265  $S_{fm}$ ? We first explore these relationships at the global scale. The long-term linear trend of the  
 266 parameters considered is variable across the globe (Figure 1 a-j2). Only  $S_{fm}$  and the parameters  
 267 that have a significant relationship with  $S_{fm}$  are presented in Figure 1. The long-term trend of  $S_{fm}$ ,  
 268 [Chl], and NPP are nuanced. These parameters are increasing at high latitudes and portions of the  
 269 subtropics;  $Z_{eu}$  is broadly the inverse of these parameters. PAR is decreasing at high latitudes and  
 270 equatorial regions and increasing in the subtropics. SST is predominately increasing over the  
 271 majority of the ocean with the exception of some regions of the central gyres and the southern tip  
 272 of South America.  $\Delta\rho_{200}$  generally follows the same spatial patters of  $Z_{eu}$  with the inverse  
 273 relationships found for MLD. SLA is increasing over the majority of the ocean.  $Q_{net}$  is primarily  
 274 variable in the equatorial region. The variability of the long-term trends of these parameters will  
 275 be explored in greater detail in the regional analysis. The correlation between these parameters  
 276 and  $S_{fm}$  is less variable (Figure 1 b-j3). Overwhelmingly, [Chl] and NPP are positively correlated  
 277 with  $S_{fm}$ . However, there are regions where [Chl] and  $S_{fm}$  are non- and anti-correlated that will be  
 278 explored in more detail in subsequent sections. Likewise,  $Z_{eu}$  is predominately negatively  
 279 correlated with  $S_{fm}$ , with the non- and anti- correlated regions inverse those of [Chl]. PAR is  
 280 generally positively correlated at high latitude and equatorial regions and anti-correlated in gyre  
 281 regions. SST is generally anti-correlated in the warm regions of the ocean and correlated at cold,  
 282 high latitude regions.  $\Delta\rho_{200}$  generally follows the same patters as SST, while MLD and  $Q_{net}$  display  
 283 an inverse relationship to SST. SLA also follows a similar correlation pattern to SST but with  
 284 weaker correlative relationships.

285 Globally, which parameters are most important to describing the variability in  $S_{fm}$ ? We  
 286 applied PLSR to the global ocean to explore this question. Light availability in the water column,  
 287 indicated as euphotic depth, is most important to the size distribution of phytoplankton, followed  
 288 by [Chl], NPP, SST, and PAR (Figure 2). Probability density plots reveal, larger cells are  
 289 associated with higher [Chl] and NPP, shallower  $Z_{eu}$ , colder SST and lower PAR. Conversely,  
 290 smaller cells are associated with deeper  $Z_{eu}$ , warmer surface waters and higher PAR (Figure S1).

291

## 292 3.2 Regional Analysis:

### 293 3.2.1 Size Relationship with Chlorophyll:

294 Are  $S_{fm}$  and [Chl] changing in synchrony? This is an important question to understand as  
 295 a subset of the satellite PFT algorithms are abundance-based, meaning they estimate PFTs directly  
 296 from empirical relationships with [Chl] (Mouw et al., 2017). Thus, the relationships used by these  
 297 approaches should only hold up where [Chl] and phytoplankton composition are strongly  
 298 correlated. It is therefore of interest to further explore regions of the ocean where strong positive  
 299 correlation between  $S_{fm}$  and [Chl] are not found. From the global analysis above, we identified  
 300 that  $S_{fm}$  and [Chl] are correlated over the majority of the ocean, but there are regions of the ocean  
 301 where there is little or no correlation between these parameters. How do the physical drivers of  
 302  $S_{fm}$  variability vary between correlated, non-correlated, and anti-correlated cases? To explore  
 303 these relationships further, we partition the ocean into regions where  $S_{fm}$  and [Chl] are correlated,  
 304 non-correlated, and anti-correlated. To isolate the impact of temperature, we further differentiate  
 305 the ocean by warm ( $\geq 18^\circ\text{C}$ ) and cold regions ( $< 18^\circ\text{C}$ ), resulting in a total of six regions (Figure  
 306 3a). We refer to these as the correlation regions.

307 In the correlated regions, only  $Z_{eu}$ , [Chl] and NPP are significant in explaining the variance  
 308 in  $S_{fm}$  (Figure 3 b, c). In these regions, [Chl],  $S_{fm}$  and NPP vary together and inversely to  $Z_{eu}$   
 309 (Figure S2). In the anti-correlated regions,  $Z_{eu}$ , [Chl] and NPP, PAR and  $Q_{net}$  are important.

310 Additionally, in the warm, anti-correlated region, MLD is also important (Figure 3 d, e). In the  
 311 cold, anti-correlated region,  $S_{fm}$ , NPP, PAR and  $z_{eu}$  vary together and inversely to [Chl],  $Q_{net}$  and  
 312 MLD, while in the warm anti-correlated region,  $S_{fm}$ , varies together with, PAR and  $z_{eu}$ , but [Chl],  
 313  $Q_{net}$ , and NPP vary together with a slight time lag (Figure S2). It should be noted that the cold  
 314 ocean anti-correlated region is very small and immediately adjacent to the anti- and un-correlated  
 315 regions found in the southern portion of the South Pacific Ocean. In the uncorrelated regions,  $z_{eu}$ ,  
 316 [Chl] and NPP are still significant with the addition of SST and  $\Delta\rho_{200}$  in the cold ocean and  $Q_{net}$  in  
 317 the warm ocean. In the uncorrelated cold ocean,  $S_{fm}$ , [Chl] and NPP are varying in opposition with  
 318 each other, while  $z_{eu}$  is the inverse of [Chl], and SST and MLD track each other identically. In the  
 319 warm uncorrelated region,  $S_{fm}$ , [Chl], NPP and  $Q_{net}$  display a similar temporal pattern offset from  
 320 each other in time and inverse to  $z_{eu}$  (Figure S2). To sum up these relationships, in all regions,  $z_{eu}$ ,  
 321 [Chl] and NPP are important and in addition, A) in the correlated regions [Chl],  $S_{fm}$  and NPP vary  
 322 together and inversely to  $z_e$ ; B) in the anti-correlated regions,  $Q_{net}$  and PAR are important factors  
 323 with addition of MLD in the cold ocean, and C) in the non-correlated regions,  $Q_{net}$  is important in  
 324 the warm ocean and SST and  $\Delta\rho_{200}$  in the cold ocean.

325

### 326 3.2.2 Physical Drivers of Phytoplankton Size:

327 For the majority of our analysis we focus on regions that were determined from a  
 328 combination of SST,  $S_{fm}$  dominance and [Chl] variability (Figure 4). We refer to these as the  $S_{fm}$   
 329 and [Chl] regions. Within these regions we ask, what are the important physical drivers of  $S_{fm}$   
 330 variability? To isolate the impact of temperature, mean SST over the time series was used to  
 331 partition the ocean into warm ( $\geq 18^\circ\text{C}$ ) and cold regions ( $< 18^\circ\text{C}$ ). Other studies have used the  $15^\circ\text{C}$   
 332 isotherm to delineate warm and cold regions (Behrenfeld et al., 2006; Siegel et al., 2013).  
 333 However, the  $18^\circ\text{C}$  isotherm corresponded better to the boundaries of  $S_{fm}$  dominance (Figure 1f1  
 334 and 4a).  $S_{fm}$  dominance was determined by assessing the percentage of time spent above the global  
 335 mean for a given pixel; microplankton were considered dominant if  $S_{fm}$  was greater than the global  
 336 mean for at least 50% of the record (Figure 1a3). The standard deviation of [Chl] ( $\sigma_{[Chl]}$ ,  $\mu\text{g L}^{-1}$ )  
 337 was used to characterize [Chl] variability. Regions were partitioned from the distribution of  $\sigma_{[Chl]}$   
 338 as greater than the 75<sup>th</sup> percentile (high variability), between the 25<sup>th</sup> and 75<sup>th</sup> percentiles (moderate  
 339 variability), and less than the 25<sup>th</sup> percentile (low variability). This results in the possibility of  
 340 twelve regions. However, not all combinations contained enough pixels to proceed with analysis.  
 341 We proceeded with nine regions (Figure 4a). These included the low and moderate  $\sigma_{[Chl]}$   
 342 percentiles ( $< 25^{\text{th}}$  and  $25^{\text{th}} - 75^{\text{th}}$ ) when the phytoplankton community was dominated by small  
 343 cells for both the warm and cold ocean, the moderate and high  $\sigma_{[Chl]}$  percentiles ( $25^{\text{th}} - 75^{\text{th}}$  and  
 344  $> 75^{\text{th}}$ ) when the phytoplankton community was dominated by large cells for both the warm and  
 345 cold ocean, and the warm, large-dominated low  $\sigma_{[Chl]}$  percentiles ( $< 25^{\text{th}}$ ).

346 PLSR was run on all nine of the  $S_{fm}$  and [Chl] regions (Figure 4). To help simplify the  
 347 variability of the primary drivers of  $S_{fm}$  across these regions as determined from the PLSR, we  
 348 have color coded a map of the regions by the dominant physical drivers (Figure 5), which are  
 349 referred to as environmental regions, and to view the importance of the parameters driving  $S_{fm}$   
 350 variability in each of these regions, we have mapped the VIP scores for each parameter considered  
 351 (Figure 6). The six environmental regions represented in figure 5 correspond to the nine  $S_{fm}$  and  
 352 [Chl] regions in Figure 4 as +SST, MLD,  $\Delta\rho_{200} - \text{NPP} = \text{small}$ , cold, moderate  $\sigma_{[Chl]}$ ; +SST and  
 353 MLD = large, warm, low  $\sigma_{[Chl]}$ ; +SLA = large, warm, moderate  $\sigma_{[Chl]}$ ; + $Q_{net}$  and MLD = small,  
 354 warm, moderate  $\sigma_{[Chl]}$ ; + $Q_{net}$ , MLD, PAR -  $z_{eu} = \text{small}$ , cold, low  $\sigma_{[Chl]}$ ; and  $z_{eu}$ , [Chl] and NPP =

355 4  $S_{\text{fm}}$  and [Chl] regions including 1) large, cold, high  $\sigma_{[\text{Chl}]}$ , 2) large, cold, moderate  $\sigma_{[\text{Chl}]}$ , 3) large,  
 356 warm, high  $\sigma_{[\text{Chl}]}$ , and 4) small, warm, low  $\sigma_{[\text{Chl}]}$ .

357 For the majority of the high latitude ocean and the central gyres,  $S_{\text{fm}}$  variability is well  
 358 explained by only variability in  $z_{\text{eu}}$ , [Chl], and NPP (Figures 4d, 4e, 4g, 4h, and 5). The importance  
 359 of  $z_{\text{eu}}$  points to light availability in the water column being the most important factor in describing  
 360 the variability of phytoplankton community size composition in these regions. These three  
 361 parameters have importance across the majority of the ocean, with  $z_{\text{eu}}$  and [Chl] having the greatest  
 362 and NPP the least importance at high latitudes (Figure 6). The four regions where only  $z_{\text{eu}}$ , [Chl]  
 363 and NPP are significant in explaining  $S_{\text{fm}}$  variance include, 1-2) cold, large-dominated, high  
 364 (Figure 4d) and moderate  $\sigma_{[\text{Chl}]}$  (Figure 4e); 3) warm, small-dominated, low  $\sigma_{[\text{Chl}]}$  (Figure 4g) and  
 365 4) warm, large dominated, high  $\sigma_{[\text{Chl}]}$  (Figure 4h). In these regions,  $S_{\text{fm}}$ , [Chl], and NPP varied  
 366 together and inversely with  $z_{\text{eu}}$  (Figure S3c, d, f,g). These regions represent the extremes of size  
 367 and chlorophyll variability: large-dominated with the greatest [Chl] variability in the warm and  
 368 cold ocean and small-dominated with the lowest [Chl] variability in the warm ocean. (The other  
 369 region that small-dominated with the lowest [Chl] variability found in the cold ocean, is more  
 370 complex and discussed later.) The only regions where one of these parameters is not significant  
 371 in describing  $S_{\text{fm}}$  variability are found in a small section of the Southern Ocean (Figure 5). NPP  
 372 is excluded from the cold, small dominated, moderate  $\sigma_{[\text{Chl}]}$  region where instead just  $z_{\text{eu}}$  and [Chl],  
 373 in addition to  $\Delta\rho_{200}$ , SST and MLD are important (Figure 4b). In this region, SST and  $\Delta\rho_{200}$  vary  
 374 together but inversely from MLD, while [Chl] and  $z_{\text{eu}}$  vary inversely and  $S_{\text{fm}}$  and [Chl] are  
 375 uncorrelated (Figure S3a). Euphotic depth is non-significant in the cold, small dominated, low  $\sigma_{[\text{Chl}]}$   
 376 region with [Chl] and NPP, in addition to PAR,  $Q_{\text{net}}$  and MLD remaining important (Figure 4c).  
 377 Here,  $S_{\text{fm}}$ , NPP and PAR vary together and inversely to [Chl], MLD and  $Q_{\text{net}}$  (Figure S3b). Overall,  
 378 PAR and  $\Delta\rho_{200}$  are only important in these Southern Ocean regions as well (Figure 6).

379 All other regions have a balance of the importance of light (indicated by  $z_{\text{eu}}$  and/or PAR)  
 380 and a mode of nutrient delivery to the surface ocean (MLD,  $Q_{\text{net}}$ , SLA and  $\Delta\rho_{200}$ ) beyond  $z_{\text{eu}}$ , [Chl]  
 381 and NPP alone (Figure 5). The VIP scores of all other variables are much lower than the ones for  
 382  $z_{\text{eu}}$ , [Chl] and NPP (Figure 6). These include the upwelling and transition regions (adjacent to the  
 383 gyres or subpolar regions). The only region where one physical parameter is significant, in  
 384 addition to  $z_{\text{eu}}$ , [Chl] and NPP, is the warm, large-dominated, moderate  $\sigma_{[\text{Chl}]}$  region, where SLA  
 385 is significant (Figure 4i). SLA varies with  $S_{\text{fm}}$ , [Chl] and NPP and inversely to  $z_{\text{eu}}$  (Figure S3h).  
 386 This region is found across the equatorial Pacific and Atlantic indicating a connection to El Niño  
 387 dynamics, regions of western boundary currents, and fringing some sub-polar regions (Figure 4,  
 388 5, and 6). SLA has some of the lowest VIP scores of all parameters (Figure 6). The warm, large-  
 389 dominated, low  $\sigma_{[\text{Chl}]}$  region is also found in the equatorial Pacific (Figure 4j and 5). Here MLD  
 390 and SST are significant in addition to  $z_{\text{eu}}$ , [Chl] and NPP.  $S_{\text{fm}}$  variability here is driven by the  
 391 deepening of MLD, leading to cooling SST, associated with equatorial counter-current dynamics  
 392 (Figure 4j) that drives the seasonal timing of  $S_{\text{fm}}$ , [Chl] and NPP peaks (Figure S3i). The mixed  
 393 layer depth remains an important driver in the remaining region (warm, small-dominated,  
 394 moderate  $\sigma_{[\text{Chl}]}$  region) in addition to  $Q_{\text{net}}$  (Figure 4f), which is found around the outer edges of the  
 395 gyres (Figures 4 and 5). Here  $Q_{\text{net}}$  leads the seasonal succession of MLD,  $S_{\text{fm}}$ , [Chl], NPP, with  
 396  $z_{\text{eu}}$  varying inversely (Figure S3f).

397  
 398 3.3 Temporal Variability:

399 At the regional scale, which parameters show the greatest change over the satellite record?  
 400 To answer this, we considered how parameters change over time (Figure 1) within the regions

401 defined in Figure 4a (Figure 7). On average, [Chl] and NPP show similar trends across the various  
 402 region, increasing nearly everywhere, except the warm, small-dominated, moderate  $\sigma_{[\text{Chl}]}$  region  
 403 that captures the equatorial counter current dynamics and additionally for NPP the warm, small-  
 404 dominated, low  $\sigma_{[\text{Chl}]}$  region which covers the central gyres. Euphotic depth shows the inverse  
 405 relationship with NPP.  $Q_{\text{net}}$  and PAR are mostly neutral across the globe. For PAR the only  
 406 exceptions are found in the Southern Ocean where the cold, small-dominated, low  $\sigma_{[\text{Chl}]}$  region show  
 407 increases and the cold, small-dominated, moderate  $\sigma_{[\text{Chl}]}$  region showing decreases. These same  
 408 two regions are slightly decreasing for  $Q_{\text{net}}$  while; warm, small-dominated low and moderate  $\sigma_{[\text{Chl}]}$   
 409 region is slightly increasing. SST and SLA are increasing across the globe. MLD is increasing  
 410 across all regions in the cold ocean and slightly decreasing in the warm ocean with the exception  
 411 of the equatorial counter current region. Likewise,  $\Delta\rho_{200}$  is decreasing in the cold ocean and only  
 412 slightly decreasing in the warm ocean, with the exception of the equatorial counter current region,  
 413 which is slightly increasing. Changes in  $S_{\text{fm}}$  are more nuanced.  $S_{\text{fm}}$  is increasing in the cold ocean,  
 414 and the dynamic regions in the warm ocean (large-dominated, high and moderate  $\sigma_{[\text{Chl}]}$ ). However,  
 415  $S_{\text{fm}}$  is declining in the warm ocean where small cells dominate and  $\sigma_{[\text{Chl}]}$  is low. Merging these  
 416 aspects together, in the warm small-dominated ocean, MLD is decreasing, while  $S_{\text{fm}}$  is decreasing  
 417 but [Chl] is increasing. This suggests a shift toward greater prevalence of small cells, which are  
 418 less dependent on nutrients introduced from mixing. In the equatorial counter current region,  $S_{\text{fm}}$   
 419 and [Chl] are decreasing while MLD is increasing suggesting a possible dilution effect.

420 To provide an example of temporal changes, we selected a transect in the Pacific Ocean  
 421 (from 8°S to 22°S, along 100°W) (location displayed in Figures 1,3 and 4) that transverses regions  
 422 that are declining in  $S_{\text{fm}}$  and [Chl] in the north and increasing in both of these parameters in the  
 423 south (Figures 1a2, b2). Over the length of the transect from north to south,  $Z_{\text{eu}}$  and MLD deepen  
 424 and cooling occurs (Figures 1e2, 1i2). The transect transverses three small-dominated, warm  
 425 ocean regions across all three  $\sigma_{[\text{Chl}]}$  percentiles (Figure 4). We use Hovmöller plots of the transect  
 426 (Figure 8) to show the transition over time of these and the other parameters with statistically  
 427 significant relationships from the PLSR analysis (Figure 4). Interannual variability is evident with  
 428 an increase in both  $S_{\text{fm}}$  and [Chl] in the northern reaches of the transect in 2004 and the boundary  
 429 of smallest percent microplankton contribution and low [Chl] to the south of the transect (Figure  
 430 8a, b). To allow comparison between percent microplankton and [Chl] simultaneously, we have  
 431 coded them on a pixel-by-pixel basis, where if the given value was above the mean it was coded  
 432 “high” and conversely for “low.” In this way, we are able to visualize where size and [Chl] are  
 433 changing concurrently or oppositely. The period of time that  $S_{\text{fm}}$  and [Chl] are changing in the  
 434 same direction (either both increasing or both decreasing) declines over time. Over the timeseries,  
 435 predominately, [Chl] is remaining above the mean, but the phytoplankton community is shifting  
 436 toward smaller cells (Figure 8c). NPP, which is a function of [Chl], available light (PAR) and  
 437 photosynthetic efficiency (which in turn is temperature-dependent) increases over the timeseries  
 438 (Figures 1c2, 8d), even though the change in [Chl] is variable between north and south (Figures  
 439 1b2, 8b), PAR is neutral (Figures 1d2) and SST is cooling (Figures 1f2, 8e). Euphotic depth is  
 440 deeper to the south and is shallowing over the timeseries (Figures 1e2, 8f), particularly to the south,  
 441 which is expected with the noted increase in [Chl] in the south. The overall change in heat flux is  
 442 neutral over the transect with minor interannual variability (Figures 1j2, 8g). MLD is slightly  
 443 deepening over the timeseries (Figures 1i2, 8h) consistent with the noted SST cooling (Figures  
 444 1f2, 8e). Merging these together, at the beginning of the transect we observe conditions where  $S_{\text{fm}}$   
 445 and [Chl] are changing together (either both high or both low, relative to the mean). Over time,

446 there are greater instances of  $S_{fm}$  and [Chl] changing on the opposite directions (one is increasing  
447 while the other is decreasing and vice versa).  
448

## 449 **4 Discussion**

### 450 4.1 Importance of Light Availability

451 Satellite radiometers sample from roughly the first attenuation length of the water column  
452 ( $1/K_d$ ) (Kirk, 1994), which often is much shallower than the mixed layer depth, consequently they  
453 are not able to fully capture water column processes associated with mixing. Thus, observing the  
454 ocean from satellite biases to a portion of the water column that is most responsive to variable light  
455 availability. To some extent, the finding of euphotic depth being the most important parameter is  
456 not surprising when considering the sampling method. That being said, it should be noted that  
457 light penetration in the water column (euphotic depth) rather than absolute incident light level  
458 (PAR) is consistently the more important parameter, with PAR only playing a significant role in a  
459 small part of the Southern Ocean.

460 Cell size is also highly influenced by how pigments are packaged within the cell, known  
461 as the packaging effect (Morel & Bricaud, 1981). Small cells have little cellular material between  
462 the chloroplast and cell wall making them highly efficient absorbers, resulting in higher magnitude  
463 and more peaked absorption. With large cells, light has to penetrate more cellular material to reach  
464 the chloroplast after passing through the cell wall, resulting in muted absorption affinity and in  
465 some cases shelf-shading (see Figure 7E in Ciotti et al., 2002). The results of the primary  
466 importance of euphotic depth in predicting changes in  $S_{fm}$  is also not surprising considering these  
467 direct relationships between light and pigment packaging within various sized phytoplankton cells.  
468

### 469 4.2 Modes of Nutrient Delivery

470 While the physics of the ocean are highly dynamic and cannot be characterized by one  
471 parameter alone, our analysis treats each parameter independently but allows co-variability. Here  
472 we aim to describe the possible modes of nutrient delivery from each parameter and also where  
473 multiple parameters are coincidentally important. First, we revisit the relationship between nutrient  
474 delivery and each physical parameter. Relatively cold expressions of SST are indicative of  
475 upwelling or wind mixing strong enough to bring deep, cold, nutrient rich water to the surface.  
476 When the MLD is deeper than the nutricline, nutrient rich water is mixed into the euphotic zone,  
477 impacting phytoplankton growth and composition. SLA is indicative of major upwelling and eddy  
478 features. There is also an inverse relationship between sea surface height and nutricline depth  
479 where a negative SLA is indicative of isopycnal uplift, and a positive SLA is indicative of  
480 deepening of the thermocline and nutricline. The  $\Delta\rho_{200}$  is the difference in density between the  
481 surface and 200 m and indicative of the stability of the water column. When the ocean is highly  
482 stratified, there is reduced vertical mixing, thus a lower likelihood of the entrainment of deep,  
483 nutrient rich water into the euphotic zone. There are many fluxes that are summed to total  $Q_{net}$   
484 including net latent heat flux, net longwave radiation, net shortwave radiation and sensible heat  
485 flux. Thus  $Q_{net}$ , is indicative of the magnitude of the temperature differential between the surface  
486 ocean and the atmosphere. A higher  $Q_{net}$  represents a greater temperature differential indicative  
487 of a source of deep, cool, nutrient-rich water.

488 Many studies have pointed to the importance of stratification in controlling phytoplankton  
489 communities (Behrenfeld et al., 2006; Polovina et al., 2008). These studies suggest that with  
490 greater warming, stratification will increase, resulting in a decrease in overall production and a

491 shift toward smaller cells. However, this simple explanation cannot be expected to work across  
492 the whole ocean and other studies have concluded interannual stratification variability is not large  
493 enough to drive a coherent phytoplankton response (Dave & Lozier, 2010). The relationships  
494 between phytoplankton response and physical drivers are complicated and depend on a host of  
495 variables (Lozier et al., 2011; Barton et al., 2014).

496 Studies that have considered a host of environmental variables have concluded differing  
497 mechanisms of phytoplankton control. Using a global-scale dataset including [Chl], PAR,  
498 nutrients, MLD, SST, latitude, longitude and month of the year, Irwin and Finkel (2008) found  
499 that SST was the best parameter for explaining [Chl] variability (51%) and that light and nitrate  
500 concentration explain 47% of the variation in [Chl]. Further, MLD, surface nitrate, SST, latitude  
501 and longitude explain 83% of this variation, thus concluding that either light or macronutrients are  
502 often limiting. Thus, most of the variation in [Chl] is explained by bottom-up mechanisms.  
503 Conversely, when examining a much longer data record for just the North Atlantic, Barton et al.  
504 (2014) found strong seasonal relationships in the physical drivers of various phytoplankton  
505 communities, but on interannual to multidecadal timescales, the links to physical changes were  
506 much weaker. They attribute this to the year-to-year variability in phytoplankton assemblages  
507 being greater than that of the physical drivers, suggesting that top-down controls and/or perhaps  
508 changes in ocean circulation may be more important than the physical parameters that they  
509 considered (wind speed, heat flux, turbulent kinetic energy generation, SST, stratification, and  
510 MLD). Wilson & Coles (2005) conducted a global analysis of [Chl], SST, MLD, thermocline  
511 depth, and nutricline depth and identified three regional mechanisms of nutrient delivery to the  
512 euphotic zone that were controlling phytoplankton distributions. These include, 1) dynamic  
513 thermocline uplift in the tropics, 2) nutrient entrainment at mid-latitudes, and 3) seasonal light  
514 limitation at high latitudes. Here we expand this debate to include consideration of satellite-  
515 estimated phytoplankton size structure at global scales. We find that our results generally follow  
516 the Wilson and Coles (2005) framework, but our regions are less latitudinally bound.

517 Beginning with the equatorial regions, we find these regions are driven by SST and MLD  
518 or SLA. SLA was only found to be significant in the large-dominated, moderate  $\sigma_{[Chl]}$ , warm ocean  
519 which is found in regions primarily impacted by upwelling and El Niño dynamics. Previous  
520 studies have documented [Chl] variability in the eastern subtropical Atlantic to be highly correlated  
521 to sea surface height variability caused by divergent surface currents leading to vertical upwelling  
522 and a shallower thermocline and nutricline (Pastor et al., 2013). Negative SLA (found during La  
523 Niña) results in higher [Chl] due to isopycnal uplift, and positive SLA (found during El Niño)  
524 results in lower [Chl] due to deepening of the thermocline and nutricline. These dynamics resulting  
525 in off-equator [Chl] variability was also described by Wilson and Adamec (2001). [Chl] and  $S_{fm}$   
526 are strongly positively correlated in this region, thus driven by the same mechanisms. SST and  
527 MLD are the driving parameters in the equatorial counter current region that is large-dominated  
528 with low [Chl] variability. Here SST and MLD are proxies to the vertical advection of nutrients  
529 to the surface ocean with the dominant nutrient delivery mode in this region (Pastor et al., 2013).  
530 SST is indicative of cool, nutrient rich water being brought to the surface and MLD is important  
531 due to the variability imposed by El Niño.

532 Moving to the subtropical ocean, Signorini et al. (2015) investigated the physical drivers  
533 of the change in [Chl] and NPP in the subtropical gyres from a satellite perspective. They found  
534 downward trends in NPP for all gyres and a downward trend in [Chl] for all gyres except the South  
535 Pacific which had a non-significant weak upward trend. They found seasonality in [Chl] was  
536 tightly coupled with variability in the MLD confirming vertical mixing is the major driver of

537 phytoplankton photosynthesis in the gyres. In our study, the subtropical gyres correspond to the  
 538 warm, small-dominated, low  $\sigma_{[\text{Chl}]}$  region where our results indicate that  $S_{\text{fm}}$  is driven by  $z_{\text{eu}}$ , [Chl]  
 539 and NPP. Our results did not indicate the importance of MLD as in the Signorini et al. (2015)  
 540 study. However, Signorini et al. (2015) regressed each of the parameters that they considered only  
 541 against time. Their results suggesting [Chl] and MLD are tightly coupled were deduced from  
 542 comparing the trend of [Chl] over time with that of the MLD over time. We also find that MLD  
 543 is generally decreasing over time in the gyres, with the exception of the North Atlantic sub-tropical  
 544 gyre, but the [Chl] trends over time in these regions are more nuanced. Our analysis considers a  
 545 longer period of time and simultaneously compares each parameter considered to  $S_{\text{fm}}$ , therefore  
 546 reporting the statistical significance of each parameter to predicting  $S_{\text{fm}}$  rather than only  
 547 considering which parameters had statistically significant trends over time.

548 In the mid-latitudes outside of the gyres, we find a variety of different drivers of  $S_{\text{fm}}$ . There  
 549 are regions dominated by MLD, SLA, and  $Q_{\text{net}}$ . We isolated the impact of SST by dividing the  
 550 ocean into warm and cold regions. Thus, the impact of changing thermal regimes within the warm  
 551 ocean are now evidenced by heat flux which is indicative of the changing seasonal heating/cooling  
 552 patterns. In agreement with Wilson and Coles (2005), we would expect MLD to impact the  
 553 phytoplankton response due to seasonally variable mixing. SLA is found to impact the mid-  
 554 latitudes in upwelling and western boundary current regions and the dynamical impact of these  
 555 processes on phytoplankton response have been well documented (Schollart et al., 2004; Clayton  
 556 et al., 2014).

557 In agreement with other studies of high latitudes and Wilson and Coles (2005), we find  $S_{\text{fm}}$   
 558 to be driven by light availability in the water column. However, we also find SST, MLD,  $\Delta\rho_{200}$ ,  
 559  $Q_{\text{net}}$ , and PAR to be important in sub-regions of the Southern Ocean indicating phytoplankton  
 560 composition in this region is not simply light driven. Ardyna et al. (2017) also found the Southern  
 561 Ocean to be latitudinally and regionally divided. At the coarse latitudinal scale, variability is  
 562 driven by seasonally varying light availability. At the regional scale, phytoplankton variability is  
 563 driven by iron supply and local advection processes.

#### 564 4.3 Relationship between $S_{\text{fm}}$ and [Chl]

566 Overwhelmingly,  $S_{\text{fm}}$  and [Chl] are changing together driven by  $z_{\text{eu}}$  and SST in the warm  
 567 ocean and  $z_{\text{eu}}$ , SST,  $\Delta\rho_{200}$ , and MLD in the cold ocean. However, for the regions where  $S_{\text{fm}}$  and  
 568 [Chl] are not positively correlated, heat flux is an important variable. When  $S_{\text{fm}}$  and [Chl] are anti-  
 569 correlated, PAR is additionally important. Heat flux indicates a temperature differential between  
 570 the ocean and atmosphere, leading to greater cloud formation and less PAR. Heat flux is a potential  
 571 indicator of nutrient delivery to the surface ocean, as deep nutrient-rich water would be cold when  
 572 brought to the surface ocean. In the anticorrelated regions, we see [Chl] increasing when  $S_{\text{fm}}$  is  
 573 small. In the uncorrelated regions, we see a timing offset with [Chl] tending to increase prior to  
 574 an increase in cell size. Physiological compensation is a likely cause. When PAR decreases due  
 575 to a thermal gradient, leading to cloud formation and when nutrients remain replete, phytoplankton  
 576 increase their chlorophyll content to more efficiently capture light. Siegel et al. (2013) found  
 577 biomass changes dominate [Chl] at high latitudes, while physiological processes dominate [Chl]  
 578 variability in the tropical and sub-tropical regions. However, they also note [Chl] changes in  
 579 coastal and equatorial upwelling areas within the tropical and sub-tropical regions were dominated  
 580 by biomass. The majority of the anticorrelated and uncorrelated regions fell in these upwelling  
 581 dominated areas not identified by Siegel et al. (2013) to be dominated by physiological  
 582 compensation. However, their study considered annual climatology of the whole SeaWiFS

583 mission, while we investigate at monthly timescales. Thus, we are capturing physiological  
584 compensation occurring at seasonal timescales.

585

#### 586 4.4 Relation to Other Observational Methods

587 Field studies have reported mixed results as to the most important drivers of phytoplankton  
588 composition. Acevedo-Trejos et al. (2013), using the Atlantic Meridional dataset found  
589 temperature and nitrite+nitrate to be the most important variables, with light not playing a  
590 significant role in structuring the community. Conversely, Brun et al. (2015), using a globally  
591 distributed dataset, found MLD to be the most important environmental parameter followed by  
592 temperature and PAR within the mixed layer. In a review of many phytoplankton groups  
593 considered separately, Boyd et al. (2010) found that nitrogen was most important for diatoms, PAR  
594 was most important for *Phaeocystis antarctica* and picocyanobacteria, and temperature was most  
595 important for coccolithophores, nitrogen-fixers and *Prochlorococcus*. Across all of these studies,  
596 temperature is a common driving parameter. When running our analysis on the whole globe (i.e.  
597 Figure 2), we found SST to be the fourth most important parameter behind,  $z_{eu}$ , [Chl] and NPP.  
598 The field studies were not using [Chl] and NPP as environmental variables, thus our results are  
599 consistent after taking this into account. Yet, light in the water column still remains the most  
600 important, most likely due to the sampling method and packaging effects discussed above. We  
601 would have liked to include varying nutrients in our study, however, such a product does not exist  
602 that is global and varying at monthly timescales. We opted not to use monthly climatology of  
603 nutrients as this would have biased the interannual relationships. Instead, we have focused on  
604 physical drivers that are indicative of nutrient delivery.

605 Modeling studies that have the luxury of full depth resolution predict globally integrated  
606 primary production will decrease (Bopp et al. 2013) as a result of reduced supply of macro-  
607 nutrients to the euphotic zone. However, the response is not uniform across the globe, rather some  
608 regions have an increase in productivity due to reduction in light limitation due to increased  
609 stratification, and higher growth rates due to increased temperatures (Taucher & Oschlies, 2011;  
610 Dutkiewicz et al., 2013). Models also suggest geographical shifts in temperature structure will  
611 dramatically change local community composition with a shift toward greater abundance of small  
612 cells (Bopp et al., 2005; Marinov et al., 2013; Dutkiewicz et al., 2013), since they require less  
613 nutrients than larger cells. In modeling results, reduced nutrient supply was most pronounced on  
614 biomass and primary productivity at lower latitudes, with increased growth rates playing a stronger  
615 role in nutrient-rich higher latitudes (Dutkiewicz et al., 2013). Our results suggest similar  
616 latitudinal variations in productivity and community size shifts, with each increasing at high  
617 latitudes and decreasing at low latitudes.

618

#### 619 4.5 Temporal Trends

620 Many studies have been aimed at predicting how phytoplankton has changed over our  
621 observational record and how it will change in the future. What is clear is that the ocean has  
622 changed and will continue to change (Barton et al., 2016) and this change may be more rapid than  
623 estimated just a few years ago (Henson et al., 2017). There will be winners and losers with shifts  
624 in geographical temperature structure dramatically changing local phytoplankton community  
625 composition (Dutkiewicz et al., 2013). Our analysis shows regional increase and decline in the  
626 size of phytoplankton over our observational record.  $S_{fm}$  is increasing in the cold ocean, and the  
627 dynamic regions of the warm ocean where MLD is increasing. However,  $S_{fm}$  is declining in the  
628 warm ocean where small cells dominate, [Chl] is increasing but has low variability, and MLD is

629 decreasing. These changes will ultimately impact food web processes (Litchman et al., 2008;  
630 2010) and carbon export (Mouw et al., 2016).

631

## 632 **5 Conclusions**

633 Light availability in the water column is the most important parameter for the size  
634 distribution of phytoplankton as sampled from a satellite platform. As expected, larger cells are  
635 associated with higher [Chl] and NPP, shallower  $z_{eu}$ , colder SST and lower PAR. When  
636 considering the ocean by major thermal regimes (cold and warm), dominant size distribution, and  
637 [Chl] variability, for the majority of the high latitude ocean and the central gyres,  $S_{fm}$  variability is  
638 well explained by only variability in  $z_{eu}$ , [Chl], and NPP indicating light availability drives the  
639 phytoplankton community. In all other regions of the ocean there is a balance of the importance  
640 of light (indicated by  $z_{eu}$  and/or PAR) and mode of nutrient delivery to the surface ocean (MLD,  
641  $Q_{net}$ , SLA and  $\Delta\rho_{200}$ ). These results point to regionally varying phytoplankton distributions,  
642 responding to variable light and mixing regimes. For the majority of the ocean,  $S_{fm}$  and [Chl] are  
643 correlated and vary together with NPP and inversely to SST and  $z_{eu}$ . There are regions of the ocean  
644 where phytoplankton size distribution and [Chl] are not positively correlated. In these regions,  
645  $Q_{net}$  becomes important, in addition to  $z_{eu}$ , [Chl], NPP. PAR is also important in anti-correlated  
646 regions and various modes of mixing (as indicated by MLD, SST and  $\Delta\rho_{200}$ ) are important in the  
647 cold ocean.

648  $S_{fm}$  is increasing in the cold ocean, and the dynamic regions in the warm ocean (large-  
649 dominated, high/moderate  $\sigma_{[Chl]}$ ) where MLD is increasing. However,  $S_{fm}$  is declining in the warm  
650 ocean where small cells dominate,  $\sigma_{[Chl]}$  is low, [Chl] is increasing, and MLD is decreasing;  
651 suggesting a shift toward greater prevalence of small cells, which are less dependent on nutrients  
652 introduced from mixing. In the equatorial counter current region,  $S_{fm}$  and [Chl] are decreasing  
653 while MLD is increasing, suggesting a possible dilution effect. Temporal change suggests the  
654 vulnerability of phytoplankton size distributions in a changing ocean will be regionally varying,  
655 ultimately impacting carbon export and food web processes.

656

## 657 **Acknowledgments and Data**

658 Funding for this effort was provided by NASA grants NNX13AC34G and NNX16AR47G. This  
659 manuscript was improved from the comments of two anonymous reviewers. We gratefully thank  
660 all of the data providers including: European Space Agency for the Ocean Colour Climate Change  
661 Initiative dataset (<http://www.esa-oceancolour-cci.org/>), NASA for providing the ocean color  
662 products that were utilized by the Ocean Colour Climate Change Initiative dataset and the  
663 photosynthetically active radiation imagery product (<https://oceancolor.gsfc.nasa.gov/>),  
664 NCEP/NCAR for hosting the reanalysis products  
665 (<https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.surfaceflux.html>),  
666 AVISO for the sea level anomaly product (<https://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/msla-mean-climatology.html#c10358>), Group for High Resolution  
667 SST for hosting the SST product (<https://data.nodc.noaa.gov/ghrsst/>), the University of Maryland  
668 for hosting the Simple Ocean Data Assimilation products (<http://www.atmos.umd.edu/~ocean/>),  
669 and Oregon State University for hosting the net productivity imagery

671 (<http://www.science.oregonstate.edu/ocean.productivity/index.php>). The percent microplankton  
672 imagery can be accessed on PANGAEA: <https://doi.pangaea.de/10.1594/PANGAEA.892211>.

673 **References**

- 674 Acevedo-Trejos, E., Brandt, G., Merico, A., & Smith, S. L. (2013). Biogeographical patterns of  
675 phytoplankton community size structure in the oceans. *Global Ecology and*  
676 *Biogeography*, 22(9), 1060–1070. <http://doi.org/10.1111/geb.12071>
- 677 Alvain, S., Le Quéré, C., Bopp, L., Racault, M.-F., Beaugrand, G., Dessailly, D., & Buitenhuis,  
678 E. T. (2013). Rapid climatic driven shifts of diatoms at high latitudes. *Remote Sensing of*  
679 *Environment*, 132, 195–201. <http://doi.org/10.1016/j.rse.2013.01.014>
- 680 Ardyna, M., Claustre, H., Sallée, J.-B., D'Ovidio, F., Gentili, B., van Dijken, G., et al. (2017).  
681 Delineating environmental control of phytoplankton biomass and phenology in the  
682 Southern Ocean. *Geophysical Research Letters*, 44(10), 5016–5024.  
683 <http://doi.org/10.1002/2016GL072428>
- 684 Barton, A. D., Irwin, A. J., Finkel, Z. V., & Stock, C. A. (2016). Anthropogenic climate change  
685 drives shift and shuffle in North Atlantic phytoplankton communities. *Proceedings of the*  
686 *National Academy of Sciences*, 113(11), 2964–2969.  
687 <http://doi.org/10.1073/pnas.1519080113>
- 688 Barton, A. D., Lozier, M.S., & Williams, R.G. (2014). Physical controls of variability in North  
689 Atlantic phytoplankton communities. *Limnology and Oceanography*, 60(1), 181–197.  
690 <http://doi.org/10.1002/lno.10011>
- 691 Behrenfeld, M., & Falkowski, P. (1997). Photosynthetic rates derived from satellite-based  
692 chlorophyll concentration. *Limnology and Oceanography*, 42(1), 1–20
- 693 Behrenfeld, M., O'Malley, R., Siegel, D., McClain, C., Sarmiento, J., Feldman, G., et al. (2006).  
694 Climate-driven trends in contemporary ocean productivity. *Nature*, 444,  
695 doi:10.1038/nature05317
- 696 Bopp, L., Aumont, O., Cadule, P., Alvain, S., & Gehlen, M. (2005). Response of diatoms  
697 distribution to global warming and potential implications: A global model study.  
698 *Geophysical Research Letters*, 32(19), L19606. <http://doi.org/10.1029/2005GL023653>
- 699 Bopp, L., Resplandy, L., Orr, J., Doney, S., Dunne, J., Gehlen, M., et al. (2013). Multiple stressors  
700 of ocean ecosystems in the 21st century: Projections with CMIP5 models. *Biogeosciences*,  
701 10, 6225 -6245. doi:10.5194/bg-10-6225-2013.
- 702 Boyd, P.W., Strzepek, R., Fu, F., & Hutchins, D.A. (2010). Environmental control of open-ocean  
703 phytoplankton groups: Now and in the future. *Limnology and Oceanography*, 55(3), 1353–  
704 1376. <http://doi.org/10.4319/lno.2010.55.3.1353>
- 705 Brewin, R.J.W., Hirata, T., Hardman-Mountford, N. J., Lavender, S. J., Sathyendranath, S., &  
706 Barlow, R. (2012). The influence of the Indian Ocean Dipole on interannual variations in  
707 phytoplankton size structure as revealed by Earth Observation. *Deep Sea Research Part II:*  
708 *Topical Studies in Oceanography*, 77-80(C), 117–127.  
709 <http://doi.org/10.1016/j.dsr2.2012.04.009>
- 710 Brun, P., Vogt, M., Payne, M. R., Gruber, N., O'Brien, C. J., Buitenhuis, E.T., et al. (2015).  
711 Ecological niches of open ocean phytoplankton taxa. *Limnology and Oceanography*, 60(3),  
712 1020–1038. <http://doi.org/10.1002/lno.10074>
- 713 Carton, J.A., Chepurin, G.A. & Chen, L. (2018). SODA3: a new ocean climate reanalysis, *Journal*  
714 *of Climate*., doi:10.1175/JCLI-D-18-0149.1
- 715 Chisholm, S.W. (1992). Phytoplankton size. *Primary productivity and biogeochemical cycles in*  
716 *the sea* (ed. by P.G. Falkowski and A.D. Woodhead), pp. 213–237. Plenum Press, New  
717 York.

- 718 Ciotti, A., & Bricaud, A. (2006). Retrievals of a size parameter for phytoplankton and spectral  
719 light absorption by colored detrital matter from water-leaving radiances at SeaWiFS  
720 channels in a continental shelf region off Brazil. *Limnology and Oceanography: Methods*,  
721 4, 237–253.
- 722 Ciotti, A., Lewis, M., & Cullen, J. (2002). Assessment of the relationship between dominant cell  
723 size in natural phytoplankton communities and the spectral shape of the absorption  
724 coefficient. *Limnology and Oceanography*, 47, 404–417. doi: 10.4319/lo.2002.47.2.0404
- 725 Clark, M., & Cramer III, R.D. (1993). The probability of chance correlation using partial least  
726 squares. *Quantitative Structure-Activity Relationships*, 12, 137-145
- 727 Clayton, S., Nagai, T., & Follows, M.J. (2014). Fine scale phytoplankton community structure  
728 across the Kuroshio Front. *Journal of Plankton Research*, 36(4), 1017–1030.  
729 <http://doi.org/10.1093/plankt/fbu020>
- 730 Dave, A.C., & Lozier, M.S. (2010). Local stratification control of marine productivity in the  
731 subtropical North Pacific. *Journal of Geophysical Research*, 115(C12), C05001–16.  
732 <http://doi.org/10.1029/2010JC006507>
- 733 Dutkiewicz, S., Scott, J.R., & Follows, M.J. (2013). Winners and losers: Ecological and  
734 biogeochemical changes in a warming ocean. *Global Biogeochemical Cycles*, 27(2), 463–  
735 477. <http://doi.org/10.1002/gbc.20042>
- 736 Finkel, Z.V. (2007). Does phytoplankton cell size matter? The evolution of modern marine food  
737 webs. *Evolution of primary producers in the Sea* (ed. by P.G. Falkowski and A.H. Knoll),  
738 pp. 333–350. Academic Press, San Diego, CA.
- 739 Finkel, Z.V., Beardall, J., Flynn, K.J., Quigg, A., Rees, T.A.V. & Raven, J.A. (2009).  
740 Phytoplankton in a changing world: cell size and elemental stoichiometry. *Journal of*  
741 *Plankton Research*, 32, 119–137.
- 742 Grant, M., Jackson, T., Chuprin, A., Sathyendranath, S., Zühlke, M., Dingle, J., Storm, T., et al.  
743 (2017). Ocean Color Climate Change Initiative Product User’s Guide Version 3.1,  
744 Plymouth Marine Laboratory, Plymouth, UK (<http://www.esa-oceancolour-cci.org>)
- 745 Gregg, W.W., & Rousseaux, C.S. (2014). Decadal trends in global pelagic ocean chlorophyll: A  
746 new assessment integrating multiple satellites, in situ data, and models. *Journal of*  
747 *Geophysical Research-Oceans*, 119(9), 5921–5933. <http://doi.org/10.1002/2014JC010158>
- 748 Guidi, L., Legendre, L., Reygondeau, G., Uitz, J., Stemann, L., & Henson, S.A. (2015). A new  
749 look at ocean carbon remineralization for estimating deepwater sequestration. *Global*  
750 *Biogeochemical Cycles*, 1–16. <http://doi.org/10.1002/2014GB005063>
- 751 Henson, S.A., Beaulieu, C., Ilyina, T., John, J. G., Long, M., Séférian, R., et al. (2017). Rapid  
752 emergence of climate change in environmental drivers of marine ecosystems. *Nature*  
753 *Communications*, 8, 14682. <http://doi.org/10.1038/ncomms14682>
- 754 Hu, C., Lee, Z. & Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans: A novel  
755 approach based on three-band reflectance difference. *Journal of Geophysical Research*,  
756 117, C01011. doi:10.1029/2011JC007395
- 757 IOCCG (2014). *Phytoplankton Functional Types from Space*. Reports of the International Ocean-  
758 Colour Coordinating Group, No. 15, IOCCG, Dartmouth, NS
- 759 Irwin, A. J., & Finkel, Z. V. (2008). Mining a sea of data: Deducing the environmental controls of  
760 ocean chlorophyll. *PLoS ONE*, 3(11), e3836. <http://doi.org/10.1371/journal.pone.0003836>
- 761 Jackson, T., & Grant, M. (2016). Ocean Colour Algorithm Blending Utilizing Fuzzy Optical Water  
762 Classification. Algorithm Theoretical Baseline Document (Ocean Colour Algorithm  
763 Blending), [http://www.esa-oceancolour-cci.org/?q=webfm\\_send/587](http://www.esa-oceancolour-cci.org/?q=webfm_send/587)

- 764 Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., et al.  
 765 (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American*  
 766 *Meteorological Society*, 77, 437-470
- 767 Kostadinov, T.S., Cabré, A., & Vedantham, H. (2017). Inter-comparison of phytoplankton  
 768 functional type phenology metrics derived from ocean color algorithms and Earth System  
 769 Models. *Remote Sensing of Environment*, 190, 162–177.  
 770 doi:<http://doi.org/10.1016/j.rse.2016.11.014>
- 771 Lee, Z.P., Carder, K.L. & Arnone, R. (2002). Deriving inherent optical properties from water  
 772 color: A multi-band quasi-analytical algorithm for optically deep waters, *Applied Optics*,  
 773 41, 5755–5772
- 774 Lee, Z.P., Du, K.P. & Arnone, R. (2005). A model for the diffuse attenuation coefficient of  
 775 downwelling irradiance. *Journal of Geophysical Research*, 110(C02016). Doi:  
 776 10.1029/2004JC002275.
- 777 Lee, Z.P., Weidemann, A., Kindle, J., Arnone, R., Carder, K.L. & Davis, C. (2007). Euphotic zone  
 778 depth: Its derivation and implication to ocean-color remote sensing, *Journal of*  
 779 *Geophysical Research*, 112, C03009. doi:10.1029/2006JC003802
- 780 Litchman, E. & Klausmeier, C.A. (2008). Trait-based community ecology of phytoplankton.  
 781 *Annual Review of Ecology, Evolution, and Systematics*, 39, 615–639.
- 782 Litchman, E., Pinto, P., Klausmeier, C.A., Thomas, M.K. & Yoshiyama, K. (2010). Linking traits  
 783 to species diversity and community structure in phytoplankton. *Hydrobiologia*, 653, 15–  
 784 28.
- 785 Lozier, M. S., Dave, A. C., Palter, J. B., Gerber, L. M., & Barber, R. T. (2011). On the relationship  
 786 between stratification and primary productivity in the North Atlantic. *Geophysical*  
 787 *Research Letters*, 38(18), L18609. <http://doi.org/10.1029/2011GL049414>
- 788 Marinov, I., Doney, S. C., Lima, I. D., Lindsay, K., Moore, J. K., & Mahowald, N. (2013). North-  
 789 South asymmetry in the modeled phytoplankton community response to climate change  
 790 over the 21st century. *Global Biogeochemical Cycles*, 27(4), 1274–1290.  
 791 <http://doi.org/10.1002/2013GB004599>
- 792 Martens, H., & Martens, M. (2000). Modified Jack-knife estimation of parameter uncertainty in  
 793 bilinear modelling by partial least squares regression (PLSR). *Food Quality and*  
 794 *Preference*, 11, 5-16. doi: [https://doi.org/10.1016/S0950-3293\(99\)00039-7](https://doi.org/10.1016/S0950-3293(99)00039-7)
- 795 McClain, C. R. (2009). A Decade of Satellite Ocean Color Observations. *Annual Review of Marine*  
 796 *Science*, 1(1), 19–42. <http://doi.org/10.1146/annurev.marine.010908.163650>
- 797 Mehmood, T., Liland, K.H. Snipe, L. & Sæbø, S. (2012). A review of variable selection methods  
 798 in Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems*,  
 799 118, 62-69. doi:10.1016/j.chemolab.2012.07.010
- 800 Mélin, F., Chuprin, A., Grant, M., Jackson, T., & Sathyendranath, S. (2016). Ocean Colour Data  
 801 Bias Correction and Merging, Algorithm Theoretical Baseline Document, doi:[http://esa-](http://esa-oceancolour-cci.org/?q=webfm_send/619)  
 802 [oceancolour-cci.org/?q=webfm\\_send/619](http://esa-oceancolour-cci.org/?q=webfm_send/619)
- 803 Morel, A., & Berthon, J.-F. (1989). “Surface pigments, algal biomass, and potential production of  
 804 the euphotic layer: relationships reinvestigation in view of remote-sensing applications.”  
 805 *Limnology and Oceanography*, (8), 1545–1562.
- 806 Morel, A., & Bricaud, A. (1981). Theoretical results concerning light absorption in a discrete  
 807 medium, and application to specific absorption of phytoplankton. *Deep Sea Research*, 11,  
 808 1375–1393. doi: 10.1016/0198-0149(81)90039-X

- 809 Mouw, C. B., Barnett, A., & McKinley, G. A. (2016). Phytoplankton size impact on export flux in  
 810 the global ocean. *Global Biogeochemical Cycles*, 30(10), 1542–1562.  
 811 <http://doi.org/10.1002/2015GB005355>
- 812 Mouw, C. B., Hardman-Mountford, N. J., Alvain, S., Bracher, A., Brewin, R. J. W., Bricaud, A.,  
 813 et al. (2017). A Consumer's Guide to Satellite Remote Sensing of Multiple Phytoplankton  
 814 Groups in the Global Ocean. *Frontiers in Marine Science*, 4, 497–19.  
 815 doi:<http://doi.org/10.3389/fmars.2017.00041>
- 816 Mouw, C. B., & Yoder, J. A. (2010). Optical determination of phytoplankton size composition  
 817 from global SeaWiFS imagery. *Journal of Geophysical Research: Oceans*, 115(C12),  
 818 C12018. <http://doi.org/10.1029/2010JC006337>
- 819 O'Reilly, J., Maritorena, S., Mitchell, B., Siegel, D., Carder, K., Garver, S., Kahru, M., & McClain,  
 820 C. (1998). Ocean color chlorophyll algorithms for SeaWiFS. *Journal of Geophysical  
 821 Research*, 103, 24937–24953
- 822 Pastor, M. V., Palter, J. B., Pelegrí, J. L., & Dunne, J. P. (2013). Physical drivers of interannual  
 823 chlorophyll variability in the eastern subtropical North Atlantic. *Journal of Geophysical  
 824 Research-Oceans*, 118(8), 3871–3886. <http://doi.org/10.1002/jgrc.20254>
- 825 Polovina, J. J., Howell, E. A., & Abecassis, M. (2008). Ocean's least productive waters are  
 826 expanding. *Geophysical Research Letters*, 35(3), L03618.  
 827 <http://doi.org/10.1029/2007GL031745>
- 828 Rousseaux, C. S., & Gregg, W. W. (2015). Recent decadal trends in global phytoplankton  
 829 composition. *Global Biogeochemical Cycles*, 29(10), 1674–1688.  
 830 <http://doi.org/10.1002/2015GB005139>
- 831 Sadeghi, A., Dinter, T., Vountas, M., Taylor, B., Altenburg-Soppa, M., & Bracher, A. (2012).  
 832 Remote sensing of coccolithophore blooms in selected oceanic regions using the  
 833 PhytoDOAS method applied to hyper-spectral satellite data. *Biogeosciences*, 9(6), 2127–  
 834 2143. <http://doi.org/10.5194/bg-9-2127-2012>
- 835 Sathyendranath, S., Grant, M., Brewin, R.J.W., Brockmann, C., Brotas, V., Chuprin, A., Doerffer,  
 836 R., et al. (2018). ESA Ocean Colour Climate Change Initiative (Ocean\_Colour\_cci):  
 837 Version 3.1 Data. Centre for Environmental Data Analysis, 04 July 2018.  
 838 doi:10.5285/9c334fbe6d424a708cf3c4cf0c6a53f5. <http://dx.doi.org/10.5285/9c334fbe6d424a708cf3c4cf0c6a53f5>.  
 839 <http://dx.doi.org/10.5285/9c334fbe6d424a708cf3c4cf0c6a53f5>.
- 840 Schollaert, S., Rossby, T., & Yoder, J. (2004). Gulf Stream cross-frontal exchange: possible  
 841 mechanisms to explain interannual variations in phytoplankton chlorophyll in the Slope  
 842 Sea during the SeaWiFS years. *Deep-Sea Research II*, 51, 173–188.
- 843 Siegel, D. A., Behrenfeld, M. J., Maritorena, S., McClain, C. R., Antoine, D., Bailey, S. W., et al.  
 844 (2013). Regional to global assessments of phytoplankton dynamics from the SeaWiFS  
 845 mission. *Remote Sensing of Environment*, 135, 77–91.  
 846 doi:<http://doi.org/10.1016/j.rse.2013.03.025>
- 847 Signorini, S. R., Franz, B. A., & McClain, C. R. (2015). Chlorophyll variability in the oligotrophic  
 848 gyres: mechanisms, seasonality and trends. *Frontiers in Marine Science*, 2, 1–11.  
 849 <http://doi.org/10.3389/fmars.2015.00001>
- 850 Soppa, M., Völker, C., & Bracher, A. (2016). Diatom Phenology in the Southern Ocean: Mean  
 851 Patterns, Trends and the Role of Climate Oscillations. *Remote Sensing*, 8(5), 420–17.  
 852 <http://doi.org/10.3390/rs8050420>

- 853 Taucher, J., & Oschlies, A. (2011). Can we predict the direction of marine primary production  
 854 change under global warming? *Geophysical Research Letters*, 38(2), L02603.  
 855 <http://doi.org/10.1029/2010GL045934>
- 856 Vantrepotte, V., & Melin, F. (2009). Temporal variability of 10-year global SeaWiFS time-series  
 857 of phytoplankton chlorophyll a concentration. *ICES Journal of Marine Science*, 66(7),  
 858 1547–1556. <http://doi.org/10.1093/icesjms/fsp107>
- 859 Wetzels, R., & Wagenmakers, E.-J. (2012). A default Bayesian hypothesis test for correlations and  
 860 partial correlations, *Psychonomic Bulletin & Review*, 19, 1057–1064. doi:10.3758/s13423-  
 861 012-0295-x
- 862 Wilson, C., & Adamec, D. (2001). Correlations between surface chlorophyll and sea surface height  
 863 in the tropical Pacific during the 1997-1999 El Niño - Southern Oscillation event. *Journal*  
 864 *of Geophysical Research*, 106(C12), 31175–31188
- 865 Wilson, C., & Coles, V. (2005). Global climatological relationships between satellite biological  
 866 and physical observations and upper ocean properties. *Journal of Geophysical Research*,  
 867 110(C10), C10001–14. <http://doi.org/10.1029/2004JC002724>
- 868 Wirtz, K. (2012). Who is eating whom? Morphology and feeding type determine the size relation  
 869 between planktonic predators and their ideal prey. *Marine Ecology Progress Series*, 445,  
 870 1–12.
- 871 Wold, S., Sjöström, M. & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics.  
 872 *Chemometrics and Intelligent Laboratory Systems*, 58, 109-130
- 873 Yoder, J.A. & Kennelly, M.A. (2003). Seasonal and ENSO variability in global ocean  
 874 phytoplankton chlorophyll derived from 4 years of SeaWiFS measurements. *Global*  
 875 *Biogeochemical Cycles*. 17(4):112. doi:10.1029/2002GB001942
- 876
- 877

878

879

**Table 1.** List of satellite imagery and reanalysis products and sources.

<b>Parameter</b>	<b>Description and source</b>	<b>Use</b>
$a_{dg}(\lambda)$ ( $m^{-1}$ )	Spectral absorption of dissolved and detrital matter Ocean Colour Climate Change Initiative (OC-CCI) <a href="http://www.esa-oceancolour-cci.org">www.esa-oceancolour-cci.org</a> QAA (Lee et al., 2002; 2007)	$S_{fm}$ calculation
[Chl] ( $\mu g L^{-1}$ )	Chlorophyll <i>a</i> concentration Ocean Colour Climate Change Initiative (OC-CCI) <a href="http://www.esa-oceancolour-cci.org">www.esa-oceancolour-cci.org</a>	$S_{fm}$ calculation & overall analysis
$K_d(490)$ ( $m^{-1}$ )	Diffuse attenuation coefficient at 490 nm Ocean Colour Climate Change Initiative (OC-CCI) <a href="http://www.esa-oceancolour-cci.org">www.esa-oceancolour-cci.org</a> (Lee et al., 2007)	$z_{eu}$ calculation
MLD (m)	Mixed layer depth Simple Ocean Data Assimilation (SODA) <a href="http://www.atmos.umd.edu/~ocean/">www.atmos.umd.edu/~ocean/</a>	Overall analysis
NPP ( $mg C m^{-2} d^{-1}$ )	Net primary production Ocean Productivity <a href="http://www.science.oregonstate.edu/ocean.productivity/">www.science.oregonstate.edu/ocean.productivity/</a> SeaWiFS and MODIS R2014 merged following Mélin (2016)	Overall analysis
PAR ( $\mu mol$ quanta $m^{-2} d^{-1}$ )	Photosynthetically active radiation NASA Ocean Color Web <a href="http://oceancolor.gsfc.nasa.gov/">oceancolor.gsfc.nasa.gov/</a> SeaWiFS and MODIS R2014 merged following Mélin (2016)	Overall analysis
$Q_{net}$ ( $W m^{-2}$ )	Net total heat flux NCEP/NCAR reanalysis <a href="http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.surfaceflux.html">www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.derived.surfaceflux.html</a>	Overall analysis
$R_{rs}(\lambda)$ ( $sr^{-1}$ )	Spectral remote sensing reflectance Ocean Colour Climate Change Initiative (OC-CCI) <a href="http://www.esa-oceancolour-cci.org">www.esa-oceancolour-cci.org</a>	$S_{fm}$ calculation
$S_{fm}$ (%)	Percent microplankton Calculated using OC-CCI $R_{rs}(\lambda)$ , [Chl], and $a_{CDM}(\lambda)$ products <a href="https://doi.pangaea.de/10.1594/PANGAEA.892211">https://doi.pangaea.de/10.1594/PANGAEA.892211</a>	Overall analysis
SLA (m)	Sea level anomaly AVISO <a href="http://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/msla-mean-climatology.html">www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/msla-mean-climatology.html</a> - c10358	Overall analysis
SST ( $^{\circ}C$ )	Sea Surface Temperature Group for High Resolution SST (GHRSSST) <a href="http://data.nodc.noaa.gov/ghrsst/">data.nodc.noaa.gov/ghrsst/</a>	Overall analysis
$z_{eu}$ (m)	Euphotic depth Calculated from OC-CCI $K_d(490)$ as $z_{eu} = 4.6/K_d(490)$ (Morel & Berthon, 1989)	Overall analysis

<b>Parameter</b>	<b>Description and source</b>	<b>Use</b>
$\Delta\rho_{200}$ (kg m <sup>-3</sup> )	Stratification index Simple Ocean Data Assimilation (SODA) Derived $\sigma_\theta$ as the difference in density between the surface and 200 m <a href="http://www.atmos.umd.edu/~ocean/">www.atmos.umd.edu/~ocean/</a>	Overall analysis
$\sigma_\theta$ (kg m <sup>-3</sup> )	Potential density Simple Ocean Data Assimilation (SODA) <a href="http://www.atmos.umd.edu/~ocean/">www.atmos.umd.edu/~ocean/</a>	$\Delta\rho_{200}$ calculation

880

881

882 **FIGURE CAPTIONS**

883

884 **Figure 1.** Phytoplankton size, represented as percent microplankton (a), and relationships with  
 885 environmental parameters, chlorophyll concentration (b), net primary production (c),  
 886 photosynthetically available radiation (d), euphotic depth (e), sea surface temperature (f), sea level  
 887 anomaly (g), stratification index (h), mixed layer depth (i), and heat flux (j). Data are  $1^\circ \times 1^\circ$ ,  
 888 monthly resolution for pixels with 100% data coverage across all variables. See Table 1 for  
 889 parameter definitions and data sources. Column 1) Mean from 1998-2015. For mean SST (f1), the  
 890  $18^\circ\text{C}$  isotherm is outlined. Column 2) Long term linear trend from the Theil-Sen approach  
 891 (significant results  $\text{BF}_{10} > 3$  shown). The transect for Hovmoller plots in Figure 8 is highlighted.  
 892 Column 3) Percentage of the  $S_{\text{fm}}$  time series above the global mean with 50% contour outlined  
 893 (a3) or Kendall rank correlation coefficient with  $S_{\text{fm}}$  (b3-j3, significant results  $\text{BF}_{10} > 3$  shown).

894

895 **Figure 2.** Global relationship of environmental variables with percent microplankton ( $S_{\text{fm}}$ ). We  
 896 used partial least square regression (PLSR) which combines predictor variables into principle  
 897 components that are then regressed with  $S_{\text{fm}}$ . The method allows co-linearity between predictors  
 898 since they all contribute to forming principle components. a) PLSR coefficients which represent  
 899 the magnitude and direction of each predictor on  $S_{\text{fm}}$  response. b) VIP scores (variable influence  
 900 on projection) represent the relative importance of each predictor to  $S_{\text{fm}}$  variability. Euphotic depth,  
 901 [Chl], NPP, SST, and PAR are the most important correlates with  $S_{\text{fm}}$ . Bars in grey are non-  
 902 significant.

903

904 **Figure 3.** Relationship between percent microplankton and chlorophyll. a) Regions for PLSR  
 905 analysis were defined by isolating the impact of temperature at the  $18^\circ\text{C}$  isotherm (Fig. 1, f1)  
 906 combined with the correlation of  $S_{\text{fm}}$  with [Chl] (Fig. 1, b3). b-g) PLSR coefficients and VIP scores  
 907 for each region. Bars in grey are non-significant.

908

909 **Figure 4.** Relationships between percent microplankton and chlorophyll variance. a) Regions for  
 910 PLSR analysis were defined by isolating the impact of temperature at the  $18^\circ\text{C}$  isotherm (Fig. 1,  
 911 f1) combined with areas dominated by small or large phytoplankton (Fig. 1a3) and variance in  
 912 [Chl] (data not shown). Variance in [Chl] was defined as regions falling greater than the 75<sup>th</sup>  
 913 percentile, between the 25<sup>th</sup> and 75<sup>th</sup> percentiles and less than the 25<sup>th</sup> percentile. The transect for  
 914 Hovmoller plots in Fig. 8 is also shown. b-j) PLSR coefficients and VIP scores for each region.  
 915 Bars in grey are non-significant.

916

917 **Figure 5.** Regions coded by significant driver of phytoplankton size variability. With few  
 918 exceptions,  $z_{\text{eu}}$ , [Chl] and NPP were important in all regions. Drivers with significant importance  
 919 beyond these three base parameters are indicated with a "+". A "-" indicates one of the base drivers  
 920 is not statistically significant for that region. The six environmental regions represented in this  
 921 figure correspond to the nine  $S_{\text{fm}}$  and [Chl] regions in Figure 4 as +SST, MLD,  $\Delta\rho_{200} - \text{NPP} =$   
 922 small, cold, moderate  $\sigma_{[\text{Chl}]}$ ; +SST and MLD = large, warm, low  $\sigma_{[\text{Chl}]}$ ; +SLA = large, warm,  
 923 moderate  $\sigma_{[\text{Chl}]}$ ; + $Q_{\text{net}}$  and MLD = small, warm, moderate  $\sigma_{[\text{Chl}]}$ ; + $Q_{\text{net}}$ , MLD, PAR –  $z_{\text{eu}} =$  small,  
 924 cold, low  $\sigma_{[\text{Chl}]}$ ; and  $z_{\text{eu}}$ , [Chl] and NPP = 4 size and [Chl] regions including 1) large, cold, high  
 925  $\sigma_{[\text{Chl}]}$ , 2) large, cold, moderate  $\sigma_{[\text{Chl}]}$ , 3) large, warm, high  $\sigma_{[\text{Chl}]}$ , and 4) small, warm, low  $\sigma_{[\text{Chl}]}$ .

926

927 **Figure 6.** Global maps of VIP scores for a) euphotic depth, b) chlorophyll concentration, c) net  
928 primary production, d) sea surface temperature, e) photosynthetically active radiation, f)  
929 stratification index, g) sea level anomaly, h) heat flux, and i) mixed layer depth. Areas shaded  
930 grey are below the VIP threshold of significance ( $VIP < 0.5$ ).

931

932 **Figure 7.** Frequency distribution of statistically significant ( $BF_{10} > 3$ ) long term linear trends from  
933 the Theil-Sen approach (Figure 1, center column) of the regions defined in Figure 4a.

934

935 **Figure 8.** Example of temporal change in percent microplankton and chlorophyll concentration.  
936 The transect traverses the South Equatorial Pacific Ocean ( $100^{\circ}W$ ) as displayed in Figure 1. The  
937 parameters depicted are those that are statistically significant with phytoplankton size. Size (a)  
938 and chlorophyll (b) is declining in the north and increasing in the south. c) To compare these  
939 parameters simultaneously, we have coded them on a pixel-by-pixel basis, where if the given value  
940 was above the mean than it was coded “high” and conversely for “low.” Net primary production  
941 (d) is increasing while SST (e) is cooling. Euphotic depth is shallowing (e). Heat flux (g) varies  
942 significantly over the annual cycle, but is predominately neutral over the whole length of the  
943 record, with a decline near the northern reaches of the transect, while mixed layer depth (h) is  
944 slightly declining across the whole transect. The white horizontal lines indicate the transitions  
945 between regions depicted in Figure 4a.

946