Phytoplankton size classes of global oceans under different bathymetric depths

Yu Huan[1], Deyong Sun[2], Shengqiang WANG[3], Hailong Zhang[4], Zhenghao Li[3], and Yijun He[5]

[1]School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing, China
[2]Nanjing University of Information Science & Technology, China
[3]School of Marine Sciences, Nanjing University of Information Science & Technology
[4]Nanjing University of Information Science & Technology
[5]Nanjing University of Information Science and Technology

November 24, 2022

Abstract

The three-component model is often used to invert the phytoplankton size class (PSC) concentration globally, especially in open oceans. Limited by the three-component model's assumption, new efforts were made to explore PSC in different water environments. Mass global cruise data sets were gathered and classified into coastal, mixed, and open ocean data sets depending on the variation in bathymetric depth. A new power three-component model was established for coastal water samples (<50 m), where the determination coefficient (R2) were 0.99, 0.51, and 0.38 for micro- (Micro), nano- (Nano), and picophytoplankton (Pico), respectively. We also updated the coefficients of the exponential three-component model in open ocean (>200 m) and found that the PSC verification results performed better in the north of -40°N oceans (R2: 0.83, 0.70, and 0.64, respectively). A smooth function for the samples in mixed ocean waters (50–200 m) was designed to obtain PSC by different weights between the power and exponential three-component models with relatively low accuracy (R2: 0.84, 0.37, and 0.14, respectively), indicative of the complex conditions in these regions. We assessed the published models' performance in coastal and open ocean samples and found an apparent underestimation of the Nano and Pico chlorophyll concentrations when their concentrations were larger than 0.2 mg m-3. The PSC proportion distribution was consistent with existing knowledge. This study evaluated the preliminary consideration of the assumption of the exponential three-component model and found that it may fail in the South Ocean, based on the global open ocean data set.

Hosted file

Phytoplankton size classes of global oceans under different bathymetric depths

Yu Huan[1], Deyong Sun[1, 2], Shengqiang Wang[1, 2], Hailong Zhang[1, 2], Zhenghao Li[1], and Yijun He[1, 2]

[1] School of Marine Sciences, Nanjing University of Information Science & Technology, Nanjing, China.

[2] Jiangsu Research Center for Ocean Survey Technology, NUIST, Nanjing, China.

Corresponding author: Deyong Sun (sundeyong@nuist.edu.cn)

Key Points:

- Modeling phytoplankton size classes under different bathymetric depths at a global scale, based on our customized three-component model
- Our model avoids the underestimation of nano- and picoplankton concentrations and enriches the previous assumption in the coastal waters
- Assessing performance of existing ten three-component models in the coastal and open ocean waters, and our established model performs best

Abstract

The three-component model is often used to invert the phytoplankton size class (PSC) concentration globally, especially in open oceans. Limited by the three-component model's assumption, new efforts were made to explore PSC in different water environments. Mass global cruise data sets were gathered and classified into coastal, mixed, and open ocean data sets depending on the variation in bathymetric depth. A new power three-component model was established for coastal water samples (<50 m), where the determination coefficient (R^2) were 0.99, 0.51, and 0.38 for micro- (Micro), nano- (Nano), and picophytoplankton (Pico), respectively. We also updated the coefficients of the exponential three-component model in open ocean (>200 m) and found that the PSC verification results performed better in the north of −40°N oceans (R^2: 0.83, 0.70, and 0.64, respectively). A smooth function for the samples in mixed ocean waters (50–200 m) was designed to obtain PSC by different weights between the power and exponential three-component models with relatively low accuracy (R^2: 0.84, 0.37, and 0.14, respectively), indicative of the complex conditions in these regions. We assessed the published models' performance in coastal and open ocean samples and found an apparent underestimation of the Nano and Pico chlorophyll concentrations when their concentrations were larger than 0.2 mg m^{-3}. The PSC proportion distribution was consistent with existing knowledge. This study evaluated the preliminary consideration of the assumption of the exponential three-component model and found that it may fail in the South Ocean, based on the global open ocean data set.

Plain Language Summary

Phytoplankton with different cell sizes has an essential effect in understanding some biochemical processes, such as the phytoplankton function groups and primary ocean productivity. The current topic had gone through some valuable researches and obtained significant achievements to some extent. The three-component model is a popular one to retrieve the phytoplankton size structure and has many different versions, which were all based on the assumption between size structure and total chlorophyll-a concentration. However, the assumption was less prominent in coastal waters and would thus cause a certain deviation. The present study aims to discuss this relationship in different water environments divided by bathymetric depth, build corresponding models to improve the deviation, evaluate existing models' performance, and analyze the underlying mechanisms for those variations. We found that the new model performs better than previous models, especially in coastal water, which avoided the underestimation of nano- and picoplankton. Also, it showed that the size structure in the South Ocean seems to be unstable. We suggest that the assumption in the three-component model should be verified first in specific waters and then inverse the phytoplankton size structure.

1 Introduction

Phytoplankton are fundamental elements of ocean water environments and contribute significantly to global primary productivity (Behrenfeld et al., 2001; Gong et al., 2003; IOCCG, 2014; Lee et al., 2015; Svendsen et al., 1995). Phytoplankton of different size structures have different rates of energy translation compared to organisms lower in the food chain and higher-level consumers (Hilligsøe et al., 2011) and show variations in their bio-optical characteristics, including their specific absorption spectra, chlorophyll concentration, and pigment package effect (Bricaud et al., 2004; Huan et al., 2021; Uitz et al., 2015; Vidussi et al., 2001; S. Wang et al., 2015). Usually, three phytoplankton size classes (PSCs) are determined, including pico- (0.2–2 μm, Pico), nano- (2–20 μm, Nano), and microphytoplankton (20–200 μm, Micro) (Sieburth et al., 1978). In recent years, diagnostic pigment analysis (DPA) has been widely used to determine the chlorophyll concentration of each size class based on in situ pigment data sets, measured by high performance liquid chromatography (HPLC) (Uitz et al., 2006; Vidussi et al., 2001). Many studies have established different PSC remote sensing algorithms based on DPA results (Brewin et al., 2014; Brewin et al., 2015; Devred et al., 2006; D. Sun et al., 2019; Uitz et al., 2006; Uitz et al., 2008).

In the context of PSC algorithms, the exponential three-component model is widely applied in different ocean waters, including the marginal seas of China (D. Sun et al., 2019; X. Sun et al., 2018), the Atlantic Ocean (Brewin et al., 2010; Brewin et al., 2014; Brewin et al., 2017; Devred et al., 2011), the Indian Ocean (Robert et al., 2012), and other global oceans (Brewin et al., 2011; Brewin et al., 2015). This model has the advantage of a concise format and making reasonable assumptions. It assumes that the total chlorophyll-a concentration (C) comprises two parts, one of which can grow to a high value with increasing C but contributes less at low C, while the other dominating the chlorophyll concentration in low C is unable to grow beyond a specific value. This assumption is first used to separate the Micro (C_{m}) and the combined Nano and Pico ($C_{\mathrm{n,p}}$) chlorophyll concentration from C and is then used to retrieve the Nano (C_{n}) and Pico (C_{p}) chlorophyll concentrations. Therefore, the applicability of this assumption is the key to this model. Many studies have successfully applied this model and achieved good inversion results (Brewin et al., 2015; Devred et al., 2011; Lin et al., 2014). However, D. Sun et al. (2019) found that the assumption could not be satisfied within $C_{\mathrm{n,p}}$ and C_{p} against C in the marginal seas of China. Furthermore, previous studies had underestimated chlorophyll concentrations of C_{n} and C_{p} using the exponential three-component model, especially C_{p} in C>0.2 mg m^{-3}. Brewin et al. (2017) modified the model to account for the influence of sea surface temperature and achieved a certain improvement. Therefore, there is a need to focus on the assumptions of this model.

The primary assumption stated above seems to be more highlighted in open ocean water than in coastal waters. Thus, classifying the ocean waters into several types may be an effective way to determine the differences in this assumption. However, there is currently no widely accepted method to evaluate water types as the distinction between different water bodies cannot be classified by the chlorophyll concentration or location alone (Yan et al., 2019). In a previous study, Bricaud et al. (1987) used remote sensing reflectance at 550 nm ($R_{\mathrm{rs}}(550)$) with a limit value of $R_{\mathrm{rs}}(550)$ ($R_{\mathrm{rs,lim}}(550)$) for a particular pigment concentration to differentiate between water bodies, which they classified as Case I waters when $R_{\mathrm{rs}}(550) < R_{\mathrm{rs,lim}}(550)$ and Case II waters when $R_{\mathrm{rs}}(550) > R_{\mathrm{rs,lim}}(550)$. However, although this method worked in open ocean waters (Case I), it was not suitable for turbid waters (Case II) with highly colored dissolved organic matter (CDOM). In a subsequent study, Zhang et al. (2005) developed a convenient method to classify Case I and Case II waters based on the ratio of $R_{\mathrm{rs}}(510)$ and $R_{\mathrm{rs}}(412)$, where waters were classified Case I when $R_{\mathrm{rs}}(510)/R_{\mathrm{rs}}(412) < 1.5$ and Case II when $R_{\mathrm{rs}}(510)/R_{\mathrm{rs}}(412) > 1.5$. Although the R_{rs} method facilitates the classification of Case I and Case II waters, in situ measured R_{rs} data sets (visible light bands, 400–700 nm) obtained by cruises are scarce and would be fewer still after matching the pigment data sets, especially for global ocean cruises. Thus, there is a need for another method to approximate the different water types for further PSC model analysis.

Case I waters are usually distributed in deep bathymetric depth, while Case II waters are located in coastal areas, where the water depth would be shallower. Therefore, bathymetric depth may be a potential index to identify different water types. The continental shelf is defined as flat land extending from the subtidal line to the edge, whose slope gradient significantly increases, with a water depth from 50 m to 500 m (Xu et al., 1999). Shepard (1973) processed mass data sets and concluded that the mean bathymetric depth of the continental shelf edge was 130 m. There is usually a 200 m isobath in the marine map, which is regarded as the depth of the continental shelf edge. In other words, within 200 m, the continental shelf is called the shallow continental shelf, and above 200 m, it is known as the deep continental shelf (Xu et al., 1999). Kuenen (1950) simulated a swell with a wavelength of 180 m and amplitude of 6 m and found that it could not drive fine sand at 200 m. On the other hand, areas within 50 m water depth could be considered the interior part of the shallow continental shelf, with the region in the range of 50–200 m as the outer part. S. Liu et al. (2015) drew suspended particulate matter (SPM) from the East China Sea and found that the mass concentration isobath of SPM was more serried within the 50 m-isobath than in other areas, and nearly less than 0.5 mg l^{-1} in the whole water layers. This phenomenon had also been observed in the Kara Sea (Kravchishina et al., 2013), the southern North Sea (Eleveld et al., 2008), and the northwest Aegean Sea (Karageorgis et al., 2003). Therefore, the 50 m-isobath roughly represents the line of turbid coastal waters.

The present study represented coastal and open ocean waters by their bathymetric depth (D), classified as open ocean waters when D > 200 m and coastal waters when D < 50 m. The water area within 50–200 m was seen as the mixed area, wherein the inherent optical properties are controlled by phytoplankton and other SPMs (such as silt). This study is the first to treat each water environment separately, i.e., shallow coastal waters (< 30 m) (Brewin et al., 2015). We aimed to: (1) gather different sub-data sets for different water types, (2) verify the feasibility of the assumption of the exponential three-component model and create robust PSC remote sensing models for the global ocean, and (3) evaluate the PSC concentration inversion between existing models and analyze the underlying mechanisms for those variations.

2 Data and methods

2.1 In situ data set

Mass samples covering the global ocean were collected from different cruises and used to identify the pigment data using HPLC. In this study, the total chlorophyll-a concentration (*C*) and seven typical pigments, expressed as fucoxanthin (Fuco), peridinin (Per), 19’-hexanoyloxyfucoxanthin (19’-Hex), 19’-butanoyloxyfucoxanthin (19’-But), alloxanthin (Allo), chlorophyll-b and divinyl chlorophyll-b (TChlb), and zeaxanthin (Zea), were used for further analysis. The entire data set was subjected to quality control to exclude uncertainty, which may be caused by the experimental method and unknown errors, from the data itself. We used several criteria to pretreat the entire data set: (1) samples collected within the top 10 m of the water column; (2) $C \geqslant 0.001$ mg m^{-3} (Uitz et al., 2006); (3) the exclusion of samples with more than three inapparent diagnostic pigments (concentration is less than 0.001 mg m^{-3}). A total of 5372 samples were left after applying the criteria. The geographical distribution of these samples is shown in Figure 1.

ETOPO1 is a 1 arc-minute global relief model of the Earth's surface that integrates land topography and ocean bathymetry (Amante et al., 2009) and downloaded from http://www.ngdc.noaa.gov/mgg/global/. In this study, the grid-registered version of ETOPO1 was used to describe the bathymetric depth at the location of each sample using a minimum spatial (latitude and longitude) difference match-up code. First, the positive value, standing for land, was set to nan (a constant defined in MATLAB, meaning not a number), while the negative value represented the water depth. Samples with a bathymetric depth >200 m were used to compile the open ocean sample data set (Data set A, 40.1%), which represents Case I waters, while samples with a bathymetric depth <50 m were used to compile the coastal data set (Data set B, 27.7%), representing Case II waters. The remaining samples, located within 50–200 m, were used to compile a third data set, Data set C (32.2%), for the mixed water conditions controlled by inorganic and organic particulate matter.

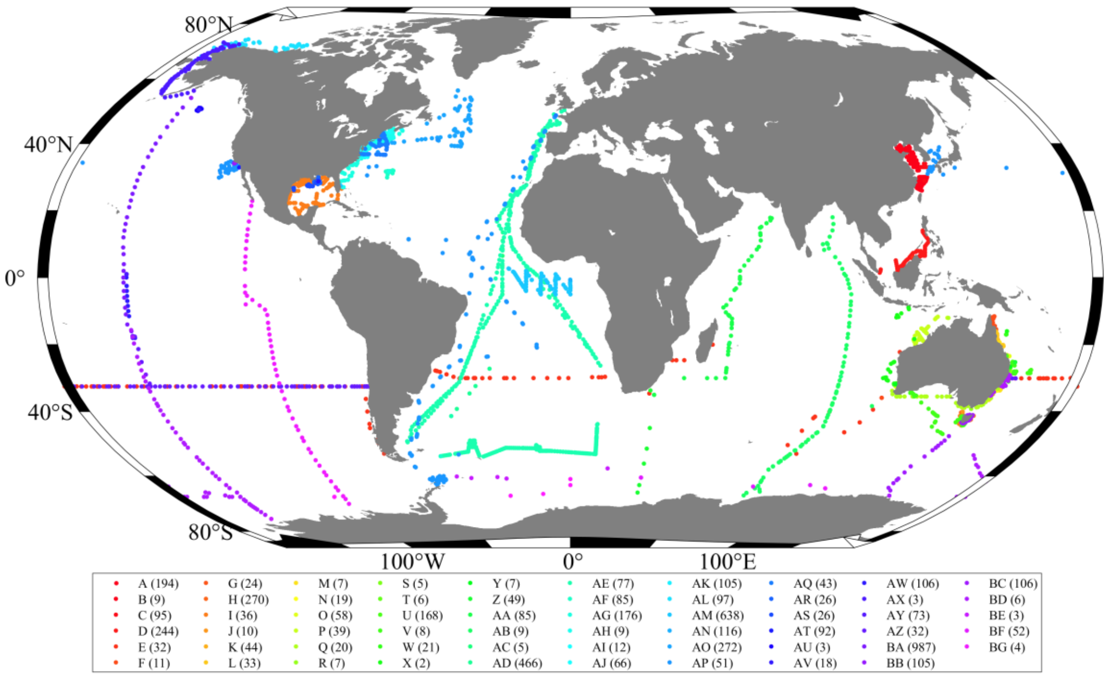

Figure 1. Locations of in situ data (≤10 m depth) used in this study. The letters denote the source of data, and the number in parenthesis denotes the number of samples. Sources of the data are as follows: A-B denotes data in the marginal seas of China (D. Sun et al., 2019) and South Korea; C denotes data in the southeastern South China Sea (Bracher et al., 2014); D-X: denotes data around the Australian (IMOS and GO-SHIP cruises, https://marlin.csiro.au/); Y-AA: denotes data in the India Ocean (including I06S_2019, I07N_2018, and i8si9n cruises); AB-AP: denotes data in the Atlantic Ocean (including ANT series, AMT28, BATS, ECOA-1, CV series, LOBO_2009, MOVE series, NASA_3Rivers, malina_admundsen_leg2, icescape2010, icescape2011, msm18-3, and naames_4 cruises, and NOMAD data set); AQ-AR:

denotes data in the Gulf of Mexico (ch-12-10 and JUN17GOM cruises); AS-BA: denotes data in the eastern Pacific Ocean [including ACIDD_2017, EXPORTSNP, KM14-16, L'Atalante (Moutin et al., 2013), ICESCAPE2011, KM14-16, p06_2017_leg2, P16N, and pb1–311 cruises]; BB-BG: denotes data around Antarctica (including BEAGLE, p16s, i89si9, io6s, NAAMES_1, p188, s04p, JR18005, KY1804, P18, PS117 cruises). The total number of samples used was 5372. The cruises without citation can be obtained on the Seabass website (https://seabass.gsfc.nasa.gov/cruise/)

2.2 Satellite data

The visible and infrared imager/radiometer suite (VIIRS) is a multi-disciplinary instrument that flows on the Joint Polar Satellite System (JPSS) series. The JPSS-1 satellite was launched in November 2017. VIIRS is designed as the successor to MODIS for Earth science data product generation, which has 22 spectral bands ranging from 412 nm to 12 µm. In this study, the annual total chlorophyll-a concentration product at a resolution of 9 km in 2020 was downloaded from the NASA Ocean Color data website (https://oceandata.sci.gsfc.nasa.gov/VIIRS-JPSS1/Mapped/Annual/9km/chl_ocx/). The total chlorophyll-a concentration was derived from the OCX algorithm (*Chl_ocx*, mg m^{-3}) (O'Reilly et al., 2000).

2.3 Phytoplankton size classes quantification

2.3.1 Diagnostic pigment analysis

Diagnostic pigment analysis (DPA) is an effective method for estimating the chlorophyll fraction of different phytoplankton size structures based on pigment data (Brewin et al., 2015; Uitz et al., 2006; Vidussi et al., 2001). In the three sub-datasets (Data sets A, B, and C), the weighted chlorophyll-a concentration, C_w, was defined as the sum of the products between seven weights of seven diagnostic pigments and corresponding concentration, shown as:

$$C_w = \sum_{i=1}^{7} W_i c_i \tag{1}$$

where W_i is the weight of the i-th pigment, and c_i is the i-th pigment concentration (the order is shown in Table 1). Initially, 2/3 samples were selected from increasing ranked samples in each sub-data set to conduct DPA, and 1/3 were used to compare the performance with that in published works. We then deduced the weights in each sub-dataset using multiple regression analysis, as listed in Table 1. The regression in different sub-datasets is highly significant, with R^2 values of 98.1%, 96.2%, and 97.4% for the open ocean, coastal, and mixed water samples, respectively ($p<0.001$). The weight of Allo is consistent with H. Liu et al. (2021), which is based on the data set in the marginal seas of China.

For deducing the chlorophyll-a fraction of the three size classes from remote sensing, the method described by Devred et al. (2011) and Brewin et al. (2015) was used because of the intersection contribution of Fuco, 19’-Hex, and 19’-But on the Micro and Nano size classes. The picophytoplankton fraction (F_p) was estimated as follows:

$$F_{\mathrm{p}} = \begin{cases} \dfrac{(-12.5C+1)W_3c_3}{C_{\mathrm{w}}} + \dfrac{\sum_{i=6}^{7} W_i c_i}{C_{\mathrm{w}}} & (\text{if } C \le 0.08 \text{ mg m}^{-3}) \\ \dfrac{\sum_{i=6}^{7} W_i c_i}{C_{\mathrm{w}}} & (\text{if } C > 0.08 \text{ mg m}^{-3}) \end{cases} \quad (2)$$

The nanophytoplankton fraction (F_{n}) was estimated as follows:

$$F_{\mathrm{n}} = \begin{cases} \dfrac{12.5CW_3c_3}{C_{\mathrm{w}}} + \dfrac{\sum_{i=4}^{5} W_i c_i + W_1 P_{1,n}}{C_{\mathrm{w}}} & (\text{if } C \le 0.08 \text{ mg m}^{-3}) \\ \dfrac{\sum_{i=3}^{5} W_i c_i + W_1 P_{1,n}}{C_{\mathrm{w}}} & (\text{if } C > 0.08 \text{ mg m}^{-3}) \end{cases} \quad (3)$$

where $c_{1,\mathrm{n}}$ refers to the part of the Fuco pigment (c_1) contributed by nanophytoplankton, which was estimated by the concentration of 19'-Hex (c_3) and 19'-But (c_4) as follows (Brewin et al., 2015; Devred et al., 2011):

$$c_{1,\mathrm{n}} = 10^{[q_1 \log_{10}(c_3) + q_2 \log_{10}(c_4)]} \quad (4)$$

where the parameters, q_1 and q_2, are 0.356 and 1.190, respectively (Devred et al., 2011; Werdell et al., 2005; Werdell et al., 2013). Lastly, the microphytoplankton fraction (F_{m}) was estimated as follows:

$$F_{\mathrm{m}} = \frac{\sum_{i=1}^{2} W_i P_i - W_1 P_{1,n}}{C_{\mathrm{w}}} \quad (5)$$

The chlorophyll concentration of each size class was obtained by multiplying C as follows:

$$\begin{aligned} C_{\mathrm{m}} &= F_{\mathrm{m}} \times C \\ C_{\mathrm{n}} &= F_{\mathrm{n}} \times C \\ C_{\mathrm{p}} &= F_{\mathrm{p}} \times C \\ C_{\mathrm{n,p}} &= C_{\mathrm{n}} + C_{\mathrm{p}} \end{aligned} \quad (6)$$

where subscripts m, n, and p represent Micro, Nano-, and Pico, respectively, and the "p,n" refers to the combined part of the Pico and Nano.

Table 1. Diagnostic pigments used as biomarkers and their taxonomic significance

Diagnostic pigments	Abbreviations	Taxonomic group	Size class	Weights of Data set			p
				A	B	C	
Fucoxanthin (c_1)	Fuco	Diatoms	M/N	1.84	1.66	1.83	<0.01
Peridinin (c_2)	Perid	Dinoflagellates	M	0.22	0.61	1.60	<0.01
19'-Hexanoyloxyfucoxanthin (c_3)	19'-Hex	Prymnesiophytes	M/N	0.66	0.44	1.31	<0.01
19'-Butanoyloxyfucoxanthin (c_4)	19'-But	Pelagophytes	N	0.62	0.75	1.42	<0.01

Alloxanthin (c_5)	Allo	Cryptophytes	N	3.16	3.97	6.24	<0.01
Total chlorophyll-b (c_6)	TChlb	Chlorophytes, Prochlorophytes	P	1.78	2.04	0.65	<0.01
Zeaxanthin (c_7)	Zea	Cyanobacteria, Prochlorophytes	P	1.23	1.36	1.15	<0.02

2.3.2 Three-component model

The exponential three-component model is based on the assumption that the total chlorophyll concentration contains two parts: one that contributes less in low chlorophyll concentration but can grow to a high value with increasing C, including Micro (C_m) and nanophytoplankton (C_n), and another that dominates low C and cannot grow above an upper limit value, including the combined nano- and picophytoplankton ($C_{n,p}$) and picophytoplankton (C_p) (Brewin et al., 2011; Brewin et al., 2014; Brewin et al., 2015; Devred et al., 2011; Sathyendranath et al., 2001; X. Sun et al., 2018). The three-component model was first used to inverse C_m and $C_{n,p}$, and then used again to divide C_n and C_p. The exponential equations are as follows:

$$\begin{aligned} C_{n,p} &= C_{n,p}^{m}\left[1-\exp\left(-S_{n,p}C\right)\right] \\ C_{p} &= C_{p}^{m}\left[1-\exp\left(-S_{p}C\right)\right] \\ C_{n} &= C_{n,p}-C_{p} \\ C_{m} &= C-C_{n,p} \end{aligned} \tag{7}$$

where $C_{n,p}{}^{m}$ and $C_{p}{}^{m}$ are the upper limit values of the combined Nano and Pico, and Pico. $S_{n,p}$ and S_p determine the chlorophyll increase in the two size classes with increasing total chlorophyll (C). In published models, the products $C_{n,p}{}^{m}\times S_{p,n}$ and $C_{p}{}^{m}\times S_{p}$ had been deduced as constants. In this study, we had defined these values (0.94 and 0.8, respectively) as reported by Brewin et al. (2015) for the same study area.

D. Sun et al. (2019) developed a power function three-component model to deduce the PSCs chlorophyll concentration for coastal waters. However, they also adopted the exponential model to inverse C_m and $C_{n,p}$ in the first step, and then used the power function as shown below:

$$\begin{aligned} C_{p} &= aC^{b} \\ C_{n} &= cC^{d} \end{aligned} \tag{8}$$

where a, b, c, and d are coefficients. A schematic representation of the primary processes used in this study is provided in Figure 2, including data pretreatment, model development, and the final satellite application.

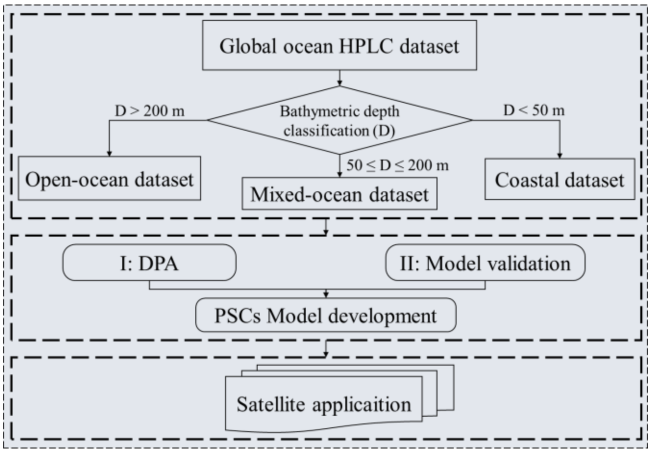

Figure 2. Schematic flow chart showing the data pretreatment, model development, and satellite application processes of phytoplankton size classes from measured HPLC data.

2.4 Performance matrix

Data processing, including data pretreatment, DPA processes, model development, and satellite application, was performed using MATLAB software (R2018b) (MathWorks Inc., Natick, MA). Several indicators were used to assess the performance of the DPA methods and the models developed in this study, including the root-mean-square error (RMSE), mean absolute percentage error (MAPE), determination coefficient (R^2), and mean ratio (MR). These metrics were calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right)^2} \tag{9}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i - y_i}{x_i}\right| \times 100\% \tag{10}$$

$$\text{Mean ratio} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{x_i}\right) \tag{11}$$

where n is the number of samples, x_{i} is the measured value, and y_{i} is the estimated value.

3 Results

3.1 Characteristics of three sub-datasets

The three sub-datasets, shown as Data sets A, B, and C for the open ocean, coastal, and mixed areas, respectively, exhibited wide variations in terms of pigment concentration. The total chlorophyll concentration (C) showed markedly different features in Data set A (Figure 3A), ranging from 0.01–10 mg m^{-3}, with a maximum frequency (96.5%) within 0.04–2 mg m^{-3}. This threshold value is the same as the upper limit of the chlorophyll-a concentration in Case I waters reported by Morel (1988). The ranges of Data set B (Figure 3C) and C (Figure 3B) were 0.07–60 and 0.04–30 mg m^{-3}, and the majority (97.4% and 99.5% for Data sets B and C) fell within a similar section, 0.1–30 mg m^{-3}.

The ratios of the seven diagnostic pigment concentrations to C were shown in Figure 3D-J. Although the three sub-datasets were determined by bathymetric depth, the distribution of the ratios revealed variable patterns in different water environments. For example, the TChlb (Figure 3I) and Zea (Figure 3J), which appeared most in the cyanobacteria cells (Kramer et al., 2019), the dominant group in Case I waters, had both the highest ratio at low *C* and a decreasing trend with increasing *C*, whereas the percentage of Fuco showed a positive trend. Fuco and Perid were often found in diatoms and dinoflagellates, often regarded as large cells and typically found in coastal waters (Bricaud et al., 2004; Devred et al., 2006; IOCCG, 2014; D. Sun et al., 2019; Uitz et al., 2006; S. Wang et al., 2015). Generally, the coverage areas of Data sets A and B are separated, while those of mixed areas (Data set C, green scatters in Figure 3D-J) tend to overlap. Some diagnostic pigment ratios of Data set C are closer to open ocean samples (such as Allo and Zea), while others are closer to the coastal samples (such as Fuco and 19'-But). Thus, despite being a rough and biased method to obtain the three sub-datasets, the biological characteristics at least partially reveal the characteristics of Case I, Case II, and mixed water environments.

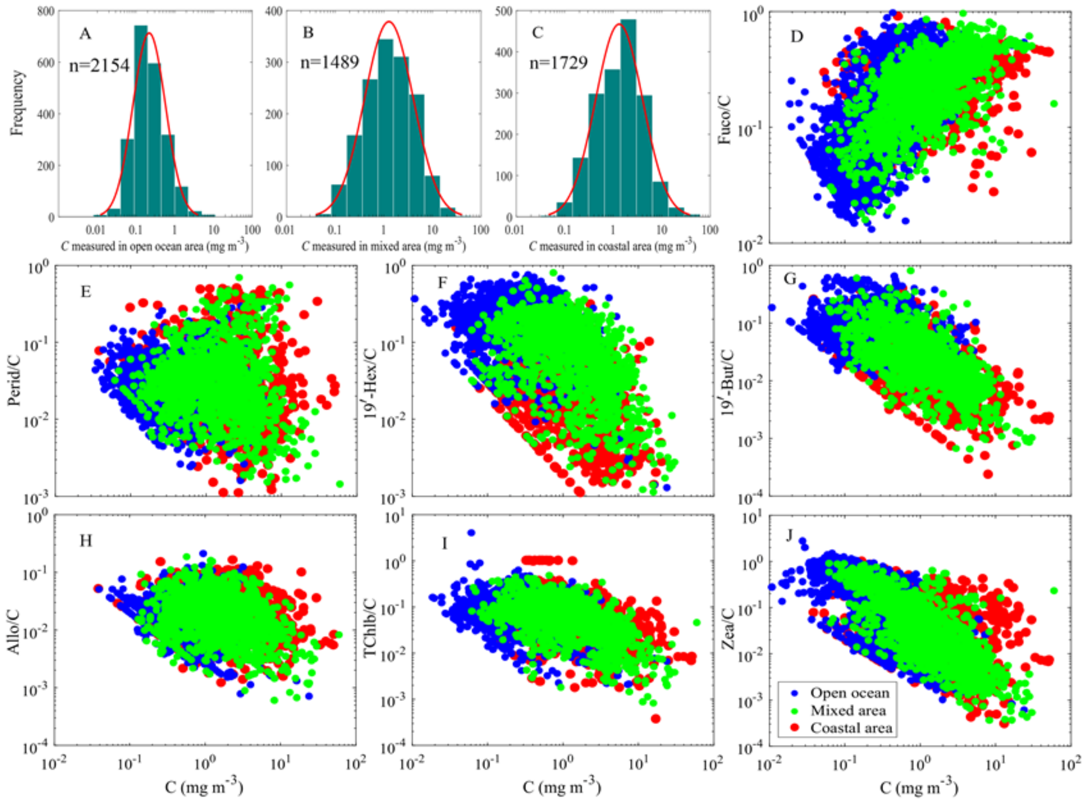

Figure 3. Pigment information of the three sub-datasets. A-C: The frequency distribution of in situ *C* in the different sub-datasets. The red line is a log-normally distributed fitting curve. D-J: Variations of the diagnosis pigments to *C* ratios as a function of *C*. The blue circle denotes the sample in open ocean waters, green in mixed areas, and red in coastal waters.

3.2 Validation results of several DPA weights

The accuracy validation of *C* and C_w was conducted as shown in Figure 4, together with the results from weights reported in five previous studies (Brewin et al., 2014; Brewin et al., 2015; D. Sun et al., 2019; Uitz et al., 2006; Uitz et al., 2008) based on the validation samples in the three sub-datasets (see section 2.3.1). The left y-axis denotes R^2 and MR, and it would be more accurate when the two indicators are closer to 1, while the lowest error of RMSE and MAPE (right y-axis) is 0. Therefore, in the open ocean samples, the DPA weights in this study showed the best performance (R^2=0.95, RMSE=0.14 mg m^{-3}, MR=1.02, and MAPE=19.15%), and the second-best was that reported by Brewin et al. (2015), which was slightly better than Brewin et al. (2014) (Figure 4A). The situation in the coastal area is nearly the same as that in Figure 4A, whereas our new DPA weights performed best (R^2=0.94, RMSE=0.17 mg m^{-3}, MR=1.02, and MAPE=19.70%) (Figure 4B). Notably, although the study by D. Sun et al. (2019) focused on the coastal water in the China marginal seas, the performance of DPA weights did not have a satisfactory result. It differed in the mixed water area, where the weights reported by D. Sun et al. (2019) had the best estimation of C_w (R^2=0.97, RMSE=0.59 mg m^{-3}, MR=1.05, and MAPE=15.5%) compared to the other studies, with the accuracy of the present study as the second-highest (R^2=0.96, RMSE=0.61 mg m^{-3}, MR=0.86, and MAPE=17.2%) (Figure 4C). Uitz et al. (2006) and Uitz et al. (2008)performed well in the open ocean and coastal waters but did not work well in mixed waters, related directly to the original data set in each study.

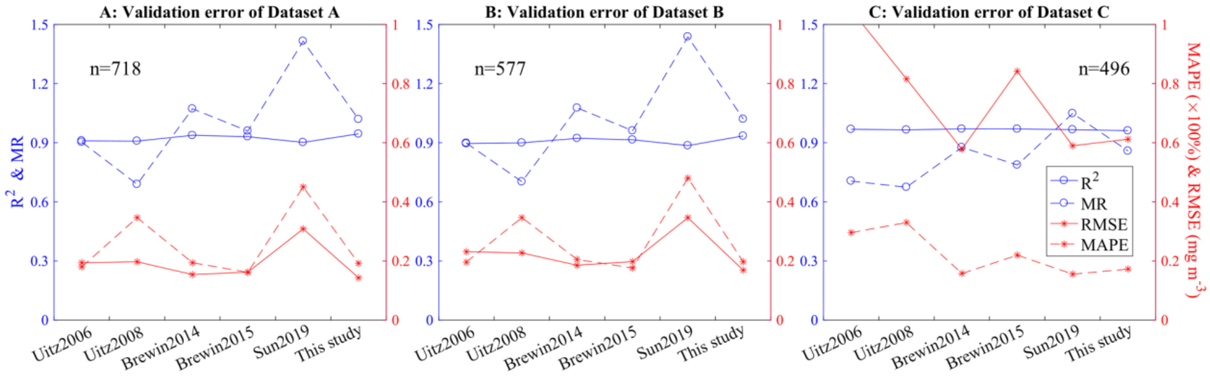

Figure 4. Error indexes of the validation between C and C_w. Different DPA weights (referred to studies in the x-axis) deduced the C_w based on Eq. 1.

3.3 Model optimization

This study aims to update and improve the remote sensing algorithms for PSCs in global oceans, especially in coastal and mixed waters. We considered DPA-deduced PSCs as the actual value to enhance the new remote sensing models. In open oceans, the three-component model is widely applied to retrieve PSCs, achieved by repeating the application of the model assumption. It assumes that the total chlorophyll concentration can be divided into two dominant parts: one that grows to a high concentration with increasing C but is not dominant at low C, and one that dominates the chlorophyll concentration at low C and is incapable of growing beyond a specific concentration (Sathyendranath et al., 2001). In the three-component model, the assumption above was applied first to divide C into C_m and $C_{n,p}$ (C_n+C_p). However, this may not be apparent when the scatters of C_p and $C_{n,p}$ were plotted as a function of C (Figure 5), especially in the coastal data set (Figure 5B). Therefore, there is an urgent need to address the effective retrieval of PSCs by further study.

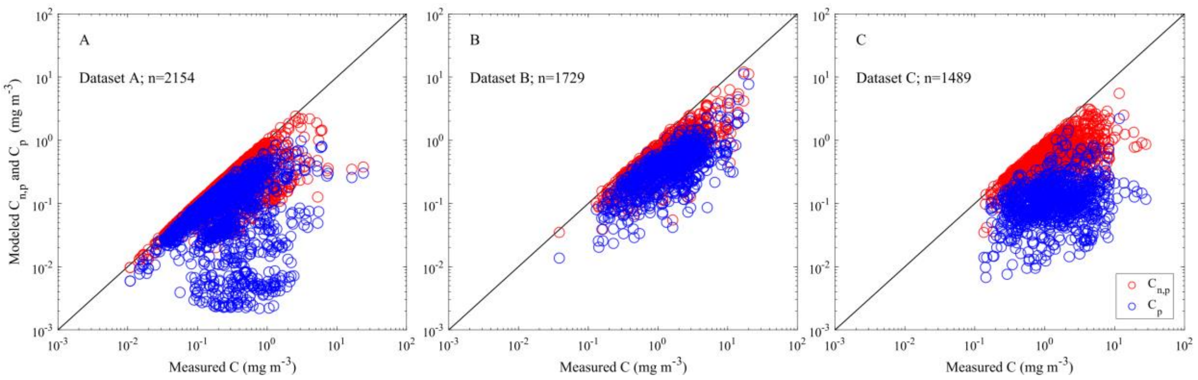

Figure 5. The chlorophyll concentrations of different size structures ($C_{\mathrm{n,p}}$ and C_{p}) as a function of total chlorophyll-a concentration (C) in different sub-datasets based on the DPA weights in this study. The black line is the 1:1 line.

3.3.1 Coastal samples

According to the $C_{\mathrm{n,p}}$ and C_{p} distributions against C in Figure 5B, these can grow to a high value with increasing C, with trends closer to a power function distribution. Thus, we designed the power function definition of $C_{\mathrm{n,p}}$ and C_{p} as follows:

$$
\begin{aligned}
&C_{\mathrm{n,p}} = aC^{b} \\
&C_{\mathrm{p}} = cC_{\mathrm{n,p}}{}^{d} \\
&C_{\mathrm{n}} = eC_{\mathrm{n,p}}{}^{f} \\
&C_{\mathrm{m}} = C - C_{\mathrm{n,p}}
\end{aligned}
\tag{12}
$$

where a, b, c, d, e, and f are the coefficients of each function and deduced by the fmincon function in MATLAB, which can find a constrained minimum of a function of several variables with a cost function (Eq. 13):

$$
\delta = \frac{\sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\left(\hat{C}_i - C_{\mathrm{PSC},i}\right)^2}}{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}} C_{\mathrm{PSC},i}}
\tag{13}
$$

where $\hat{C}_i$ and $C_{\mathrm{PSC},i}$ are the estimated and measured chlorophyll concentrations of the i-th points under the same size structures, and N is the sample number. The coefficients for a, b, c, d, e, and f are 0.434, 0.627, 0.514, 0.920, 0.388, and 1.145, respectively. We also reconstructed $C_{\mathrm{n,p}}$ as an exponential function based on the format of the two-component model, while C_{p} and C_{n} were the power functions in Eq. 12. The PSC retrievals of the two versions were shown in Figure 6. The inversion results of C_{m} both showed high consistency with the measured values for large chlorophyll concentrations. However, differences were observed in inverse ($C_{\mathrm{m}} < 1$ mg m^{-3}, approximately) (Figure 6A). The C_{m} scatters nearly all located around the 1:1 line (Figure 6A), whereas the RMSE values were 0.55 and 0.46 mg m^{-3}, and MAPE values were 54% and 65% for exponential and power functions. As expected, $C_{\mathrm{n,p}}$ from the new power function-based model had a better fit than the exponential function-based model, with more data clustered approximately 1:1 around the line, especially at high chlorophyll concentrations

(Figure 6B). The R^2 of the power function-based model was 0.54, and the RMSE was 0.46 mg m^{-3}, which was better than that of the exponential function-based model (R^2=0.38 and RMSE=0.55 mg m^{-3}). Meanwhile, the estimated C_n and C_p were also compared with the in situ values in Figure 6C-D, while the power function-based model was found to perform better than the exponential function-based model.

We also deduced the new coefficients of the exponential three-component model and retrieved the PSC concentration based on the coastal samples. The results showed good performance at low concentrations but were dispersed once these reached their upper limit. The RMSE, R^2, and MAPE values were 0.28 mg m^{-3}, 0.45, and 295% for C_n and 0.36 mg m^{-3}, 0.16, and 230% for C_p, respectively (figure not shown).

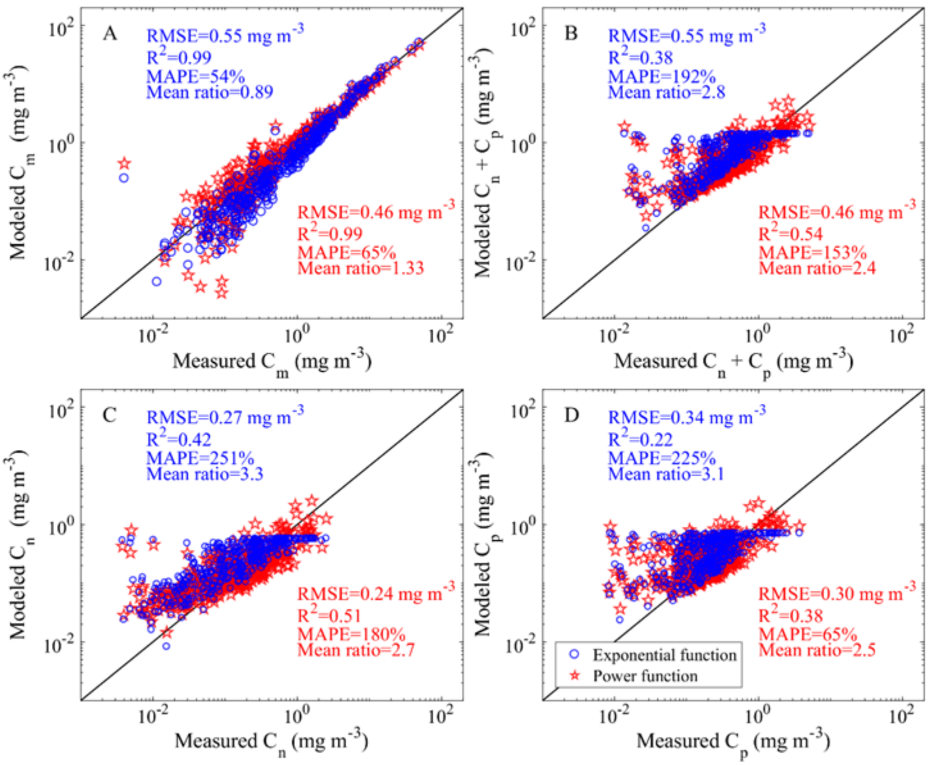

Figure 6. Comparison between modeled and measured PSCs in coastal areas based on the validation data set (n=577). Red stars are from $C_{n,p}$ based on the exponential model, and blue circles represent the PSCs from the new power function model (Eq. 12).

3.3.2 Open ocean samples

The exponential three-component model is still applicable in open ocean waters; therefore, we refitted the model coefficient and tested its accuracy (Figure 7). The PSC majority of the validation samples was found along the 1:1 lines (dashed lines in Figure 7). However, some disperse points increased the error and resulted in overestimation. C_m had the highest R^2 (0.95) and a high MAPE (334%), indicating the total deviation between the modeled and measured values (Figure 7A). The MAPE and MR in the C_p (Figure 7C) may be affected by the discrete points. If these points are removed, the error indices will perform better. On the other hand, owing to the structure of the three-component model, the upper limits of the modeled concentration in $C_{n,p}$, C_n, and C_p were still visible at high chlorophyll concentrations

(Figure 7B-C).

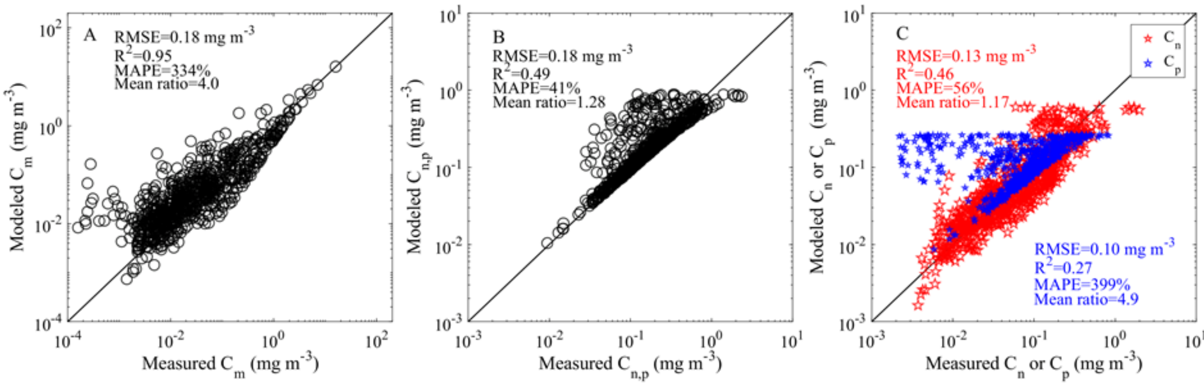

Figure 7. PSC inversions based on the three-component model in open ocean areas based on the validation data set (n=718). Red stars denote C_n, and blue stars denote C_p.

3.3.3 Model development for the intermediate region (50–200 m)

Figure 8A-D showed the retrieval of PSCs based on the new coastal and open ocean models against the validation data set of sub-dataset C (n=465). The sub-dataset C represents bathymetric depths from 50 to 200 m, which are primarily located on the continental shelf, mainly affected by turbid coastal waters and open ocean clear waters. Therefore, we assumed that the coastal and open ocean waters all had a particular contribution to the phytoplankton size structure in this area. The microphytoplankton concentration from the open ocean model performed slightly better than the coastal model, where R^2, MAPE, and MR for the former were 0.87, 42.7%, and 1.16, respectively. Similar to C_m, the open ocean model also performed slightly better than the coastal model in C_n inversion (Figure 8C). However, the coastal model held an advantage in estimating $C_{n,p}$ (Figure 8B) due to eliminating the upper limit concentration. The R^2 of $C_{n,p}$ from the open ocean and coastal models was 0.06 and 0.31 (Figure 8B), while C_p was 0.002 and 0.17 (Figure 8D), respectively. Compared with other size classes, the inversion of C_p was considerably overestimated in both models, indicating that the phytoplankton conditions may be more complex. Further study will be needed to explore this inaccuracy.

Considering the performance of the two models, we decided to use a smooth function (α and β) to retrieve the PSC concentration in the range of 50–200 m, as described by (Hu et al., 2012). In general, the bathymetric depth indicates the weight of the PSC concentration; that is, the coastal model would contribute more when the points are closer to 50 m, while the open ocean model would contribute more closer to 200 m. A detailed format of the calculations was given in Eq. 14-15, and new PSC inversion in mixed water areas was shown in Figure 8E-F, and the performance of PSCs improved to some extent. However, this represents a weighted average of the coastal and open ocean models, where the picophytoplankton remained overestimated.

$$\mathrm{PSC}=\begin{cases}\mathrm{PSC}_{\mathrm{M50}} & [\text{for } \ \textit{Water depth} \le 50\mathrm{m}] \\ \mathrm{PSC}_{\mathrm{M50\text{-}200}}=\alpha\times\mathrm{PSC}_{\mathrm{M50}}+\beta\times\mathrm{PSC}_{\mathrm{M200}} & [\text{for } \ 50<\textit{Water depth} \le 200\mathrm{m}] \\ \mathrm{PSC}_{\mathrm{M200}} & [\text{for } \ \textit{Water depth} > 200\mathrm{m}]\end{cases} \tag{14}$$

where PSC_{M50}, $PSC_{M50–200}$, and PSC_{M200} stand for the PSC model in coastal waters, mixed and open ocean area, respectively. Smooth indexes, α and β, are defined as:

$$\begin{aligned}\alpha &= (200-D)/(200-50) \\ \beta &= (D-50)/(200-50)\end{aligned} \tag{15}$$

where D is the water depth.

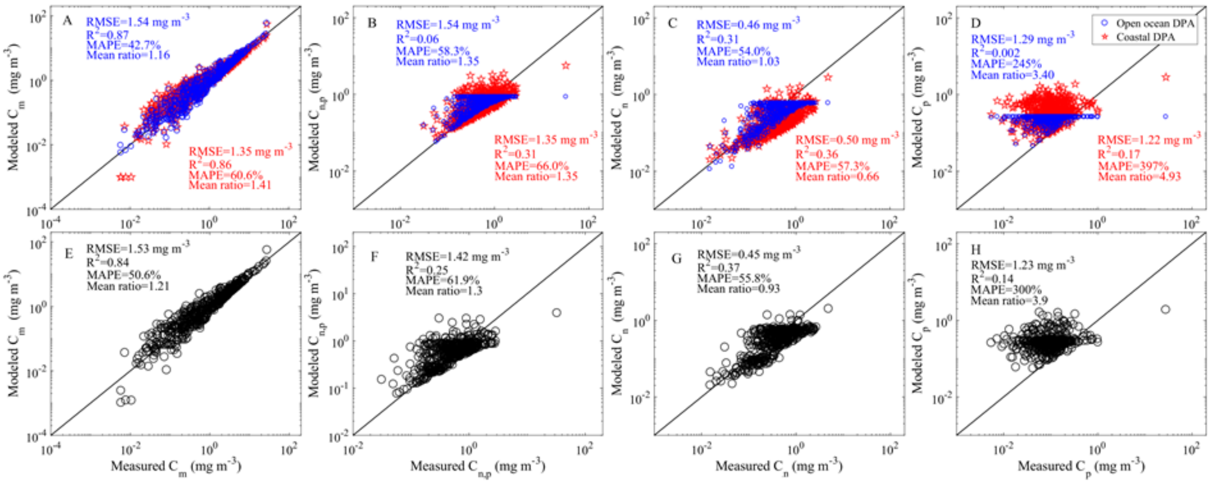

Figure 8. Results of the inversion of the PSC concentration in the validation data set of mixed waters (n=465). A-D: PSCs performance of the coastal model (red stars) and open ocean model (blue circles). E-H: PSC inversions based on Eq. 14–15.

3.4 Spatial distribution of PSCs

To visualize the distribution of the PSC proportions around the global oceans, we applied the new models to the VIIRS satellite data, as shown in Figure 9. The satellite data was the annual total chlorophyll concentration for the 2020 year. Another local region, the northwest Pacific Ocean (90°E – 180°E, 0–60°N), was also chosen for detailed information on the continental shelf (Figure 9E-H).

The general size structure spatial distribution feature agreed with that reported in the literature (Brewin et al., 2015; Hirawake et al., 2011; Mouw et al., 2010; Uitz et al., 2006), but also showed some variation. As expected, microphytoplankton (F_m) dominated the nearshore area, especially in the middle and high latitude area in the Northern Hemisphere (Figure 9A) and also appeared in the great calcite belt (-45°N), a region of elevated upper ocean calcite concentration in spring and summer in the Southern Ocean derived from coccolithophores (W. M. Balch et al., 2005; Smith et al., 2017). This zone is also known for its diatom predominance (William M. Balch et al., 2016; Rosengard et al., 2015; Smith et al., 2017). The distribution in Figure 9E showed a similar feature to that of the global oceans. However, although the bathymetric depth in the eastern region of the Kuril Islands can be up to 4,000–6,000 m, F_m still occupies approximately 50% of the phytoplankton bloom in those areas.

The combined proportion of nano- and picophytoplankton ($F_{n,p}$, Figure 9B) showed an inverse characteristic with F_m. The Nano proportion (F_n) was more stable in the whole area but was relatively higher (about 30%–40%) in mid-to-high latitudes and lower in the subtropical gyres (approximately 10%) (Figure 9C). In comparison, picoplankton (F_p) dominated the subtropical regions (60%–70%) and contributed more (50%–60%) to the total chlorophyll-a concentration in the south of –40°N than the middle and high-latitude (30%–50%) in the Northern Hemisphere (Figure 9D). The F_n in Figure 9G was slightly lower than that reported in previous studies (Brewin et al., 2015; Roy et al., 2013). G. Wang et al. (2013) reported that the F_n was approximately 35% and 10% for summer and winter in the northern South China Sea, respectively, while F_p was 60% and 80% in the same region. Compared with the distribution shown here, the results reported by G. Wang et al. (2013) are consistent with our results in Figure 9G-H.

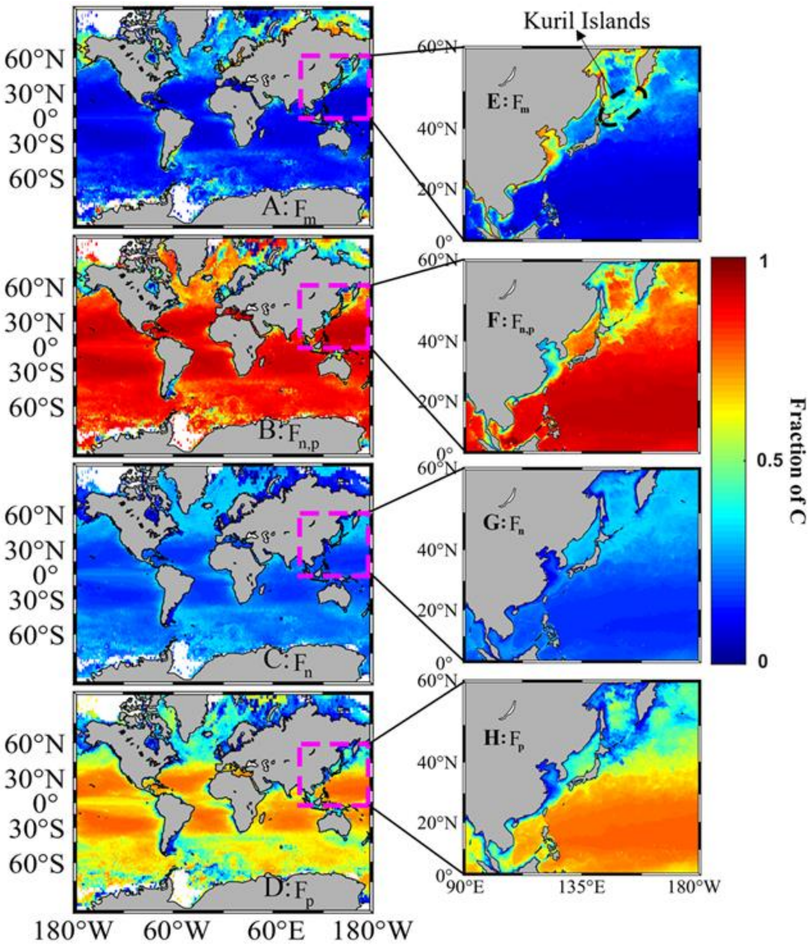

Figure 9. Annual averages of PSCs proportions (F_m, $F_{n,p}$, F_n, and F_p) in 2020 based on the new model in the global oceans (A-D), together with that in the northwest Pacific Ocean (E-H).

4 Discussion

4.1 Performance of three-component models

This study aims to assess the usability of three-component models and update the coefficients of those models based on different water types. Three-component models in the exponential format rely on assumptions that limit their inversion accuracy. In this study, we assessed the applicability of the performance of published models in coastal and open ocean data sets. In total, 10 series coefficients of the exponential three-component models were collected in this study (Table 2).

Table 2. Parameter values for Eq. 7 derived in other studies.

Study	Parameters				Location	n
	$C_{n,p}^m$	$S_{n,p}$	C_p^m	S_p		
Brewin et al. (2010)	1.060	0.849	0.110	6.636	Atlantic	1935
Brewin et al. (2011)	0.780	1.141	0.150	5.000	Global	256
Devred et al. (2011)	0.550	1.818	0.150	6.667	NW Atlantic	733
Robert et al. (2012)	0.940	1.032	0.170	4.824	Indian Ocean	712
Brotas et al. (2013)	0.360	2.556	0.070	11.000	NE Atlantic	1100
Brewin et al. (2014)	1.790	0.525	0.370	1.784	Atlantic Ocean	816
Lin et al. (2014)	0.950	0.990	0.260	3.500	South China Sea	166
Brewin et al. (2015)	0.770	1.221	0.130	6.154	Global	5841
X. Sun et al. (2018)	0.329	3.040	0.052	17.577	marginal seas of China	180
D. Sun et al. (2019)	1.692	0.591	/	/	marginal seas of China	246

The coastal data set was collected from bathymetric depths <50 m, which previous studies may not consider (Brewin et al., 2010; Brewin et al., 2015; Devred et al., 2011). These nearshore waters usually contain abundant inorganic suspended particulate matter and large phytoplankton, complicating the bio-optical conditions and thus bringing greater uncertainty to size structure inversion (D. Sun et al., 2017; D. Sun et al., 2019). The PSC fraction inversion was plotted in Figure 12 based on the coefficients in Table 2, together with their error indices in the coastal samples. Generally, the inversion of microphytoplankton chlorophyll concentration performed better than the other two size classes, indicating that the model assumption for dividing C_m and $C_{n,p}$ was effective. The coefficients reported by Brewin et al. (2014) and D. Sun et al. (2019) had nearly the same or higher accuracy in retrieving C_m than other studies, where RMSE was about 0.89 mg m^{-3}, MAPE was 80.6%, and MR was about 1.1 (Figure 10J). The error of nano- and picophytoplankton retrieval increased to a high range, especially for C_p. The mean R^2 in Figure 10K was approximately 0.24, and MAPE is almost 240%, whereas C_p is 0.73 mg m^{-3}, 0.07, and 77% for RMSE, R^2, and MAPE, respectively, shown in Figure 10L. Note that D. Sun et al. (2019) developed their coefficients based on the coastal samples in the marginal seas of China and thus had relatively better results in C_p (R^2=0.19, RMSE=0.63 mg m^{-3}, MAPE=94%). This may indicate that the method reported by D. Sun et al. (2019) is feasible for PSC chlorophyll concentration inversion in coastal waters. This method was developed in the present study and showed better results.

The exponential three-component model was much better retrieval in the open ocean data set (Figure 11). Surprisingly, D. Sun et al. (2019) reported the best C_m inversion, and the second-best was from Brewin et al. (2014). The C_n inversion showed a significant improvement in the open ocean data set (Figure 11D-F) with a mean R^2, RMSE, and MAPE of 0.49, 0.12 mg m^{-3}, and 108%, respectively. Lin et al. (2014) reported the best estimation, followed by Devred et al. (2011). Notably, the chlorophyll range of C_n was approximately 0.001–2 mg m^{-3}, which was smaller than that in the coastal area (0.005–10 mg m^{-3}) (Figure 10D-F), and the upper limit was less apparent in open ocean waters. Meanwhile, the C_p also performed better , with a mean R^2, RMSE, and MAPE of 0.24, 0.11 mg m^{-3}, and 231%, respectively. Unlike the coastal area, the low picophytoplankton chlorophyll (0.01–0.1 mg m^{-3}) was highly consistent with the values measured in open ocean waters (Figure 11G-I), close to the 1:1 line, with an upper limit that would still be evident at increasing concentrations. This phenomenon has also been reported in previous studies (Brewin et al., 2015; X. Sun et al., 2018). Therefore, the successful application of these exponential three-component models indicates the reasonableness of the primary hypothesis in clear open ocean waters.

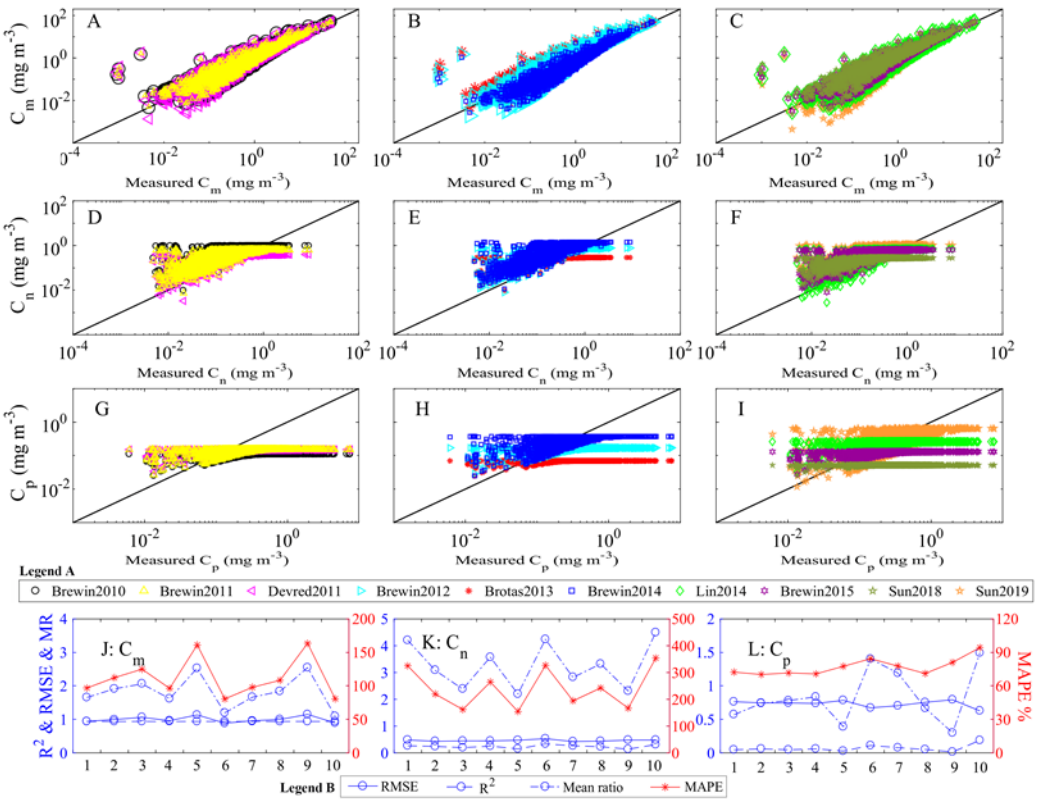

Figure 10. Modeled chlorophyll plotted against in situ chlorophyll in coastal samples, and the error indexes for each of the size fractions based on PSCs remote sensing models (three-component model) in several studies. The x-axis value in A-I is deduced from the DPA coefficients based on Data set A. The number 1–10 in the x-axis of J-L denotes the studies from left to right in Legend A.

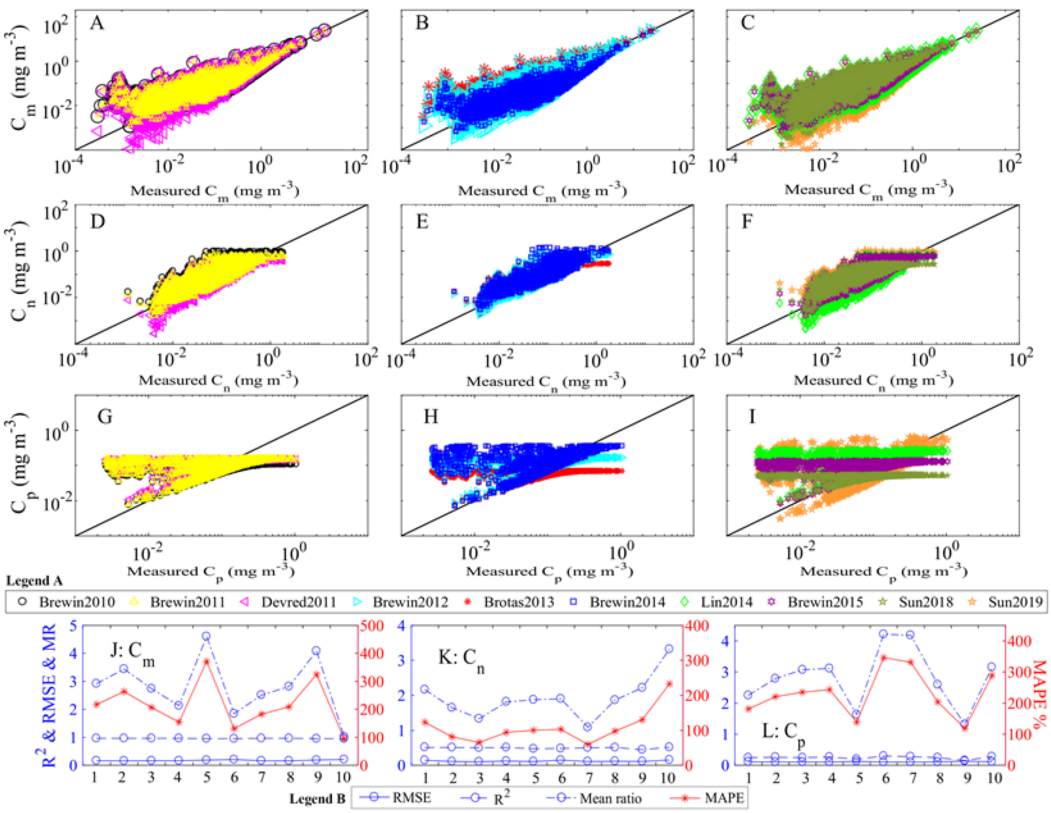

Figure 11. Modeled chlorophyll plotted against in situ chlorophyll in open ocean samples. Other information is the same as in Figure 10.

4.2. Limitation of the three-component model

The hypothesis played a vital role in the exponential three-component model. However, $C_{n,p}$ and C_p were found to increase to a wide range as C increased, as shown in Figure 5A; in other words, there was no fixed upper limit of $C_{n,p}$ ($C_{n,p}^m$) and C_p (C_p^m) for the exponential three-component models in coastal waters. For this reason, PSC retrieval would not always work, especially in $C_{n,p}$ and C_p. Thus, we further selected the samples at deeper bathymetric depth (>1000 m) from the open ocean data set (Data set A) and refined the samples in four bathymetric depth ranges, including 1000–2000 m, 2000–3000 m, 3000–4000 m, and >4000 m. As a result, there were no noticeable features that could be used to distinguish the wide $C_{n,p}$ and C_p ranges.

Next, we classified the samples in different locations. The four main ocean areas were divided approximately according to their accepted sites: the eastern Pacific Ocean (-180°E – -70°E), the Atlantic Ocean (-70°E – -15°E), the Indian Ocean (-15°E – 110°E), and the western Pacific Ocean (110°E – 180°E). Then, the north-south dividing line in each area was determined according to the $C_{n,p}$ and C_p distributions against C. The $C_{n,p}$ and C_p distributions in the Middle East–Pacific, North Atlantic, North Indian Ocean, and Northwest–Pacific were found to match the exponential three-component models well, as shown in Figure 12A-B, and the stray points in Figure 12C-D were mainly gathered in the South and Northeast Pacific. Finally, the four main ocean areas were divided into two or three sections: north, middle, and south areas, as shown in Figure 12E. As a result, we found that the boundary in the four main oceans was consistent at approximately 40°S–30°S. This phenomenon was unexpected and indicates that different pigment information needs to be further considered in the South Ocean. The new open ocean models (section 3.3.2) were applied again to the middle and north areas (not including the North Pacific area in Figure 12E), and the PSC chlorophyll concentration was plotted in Figure 12F-G. The inversion accuracy had a noticeable improvement, in particular for C_n and C_p, where RMSE, R^2, and MPAE were 0.1 mg m^{-3}, 0.7, and 47.3% for C_n and 0.06 mg m^{-3}, 0.64, and 26.2% for C_p.

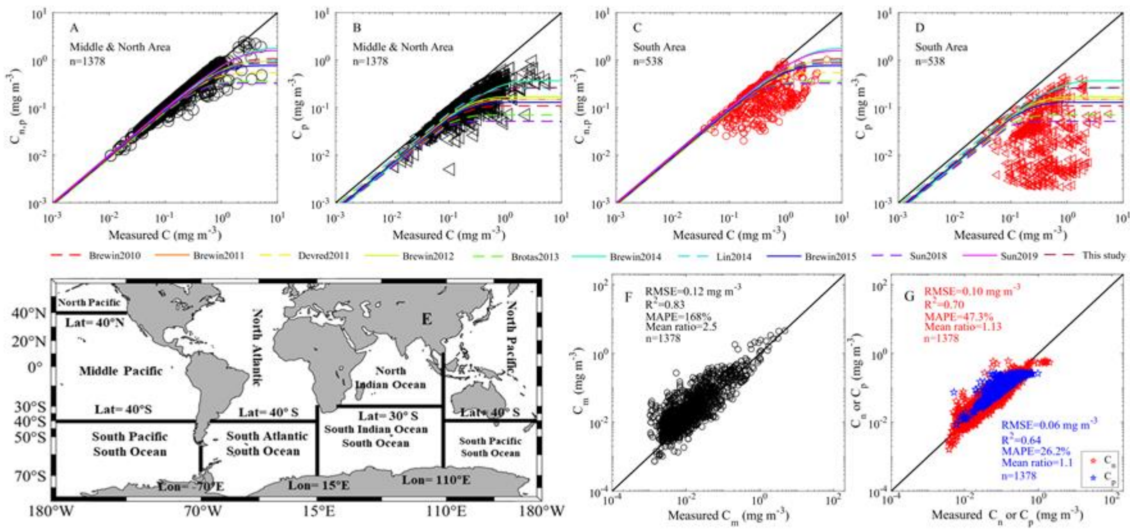

Figure 12. A-D: The combined nano- and picophytoplankton concentrations ($C_{n,p}$, A and C) and picophytoplankton chlorophyll concentration (C_p, B and D) against the total chlorophyll-a concentration in the samples at a bathymetric depth >1000 m. E: Area classification of the global oceans. F-G: PSC concentration inversion of the samples in subplots A and B.

4.3. Implications for future research work

The current exponential three-component model for PSCs remains the mainstream model for the study of phytoplankton in the world oceans. Compared with other retrieval methods, the advantage of the three-component model lies in assumptions based on in situ measured pigment data sets and its simple format. The analysis above shows that the progressive maximum value in the hypothesis ($C_{n,p}^m$ and C_p^m) may be variable in different ocean areas, as reported by previous studies (Table 2). Therefore, the mechanism by which these important parameters are distributed worldwide deserves further investigation, and the present study provides a foundation for future work in this regard. Furthermore, the accuracy of the estimation of C_p was low in mixed ocean waters and will also require further study. Although different methods were applied to deduce C_p, few improvements were obtained (data not shown).

5 Conclusions

The development of the three-component models for PSC inversion at different spatial scales could provide novel insights into the relationship between phytoplankton and ocean environment variability. The current exponential three-component model is designed for Case I waters, and further work is required to determine whether the model parameters or the primary relationship between PSC concentration and chlorophyll-a concentration remain unchanged. This study divided global ocean data sets into three parts with different pigment distribution features by bathymetric depth (D): coastal water when D < 50 m, open ocean water when D > 200 m, and mixed areas when D = 50–200 m. We analyzed the relationship in each part and built three individual models. The verification results showed that most PSC scatters were distributed along the 1:1 line, where the mean R^2 values were 0.89, 0.53, and 0.39 for Micro, Nano, and Pico, respectively. The assessment of the other ten exponential models indicated that the power three-component model performed better than the exponential model in coastal data sets. An underestimation for the Nano and Pico chlorophyll concentrations was apparent, mainly when the Pico concentration was larger than 0.2 mg m^{-3} in the open ocean samples. We further classified the DPA-derived PSC concentration distribution in the open ocean samples (D > 1000 m) and found that the Nano and Pico concentrations south of –40°N presented a highly dispersive state, especially in Pico, but consistent with the assumption in the north of −40°N. The results in mixed areas showed less stability in the Pico inversion, which may be caused by its complex water conditions. The separate discussions of the PSC models provide a comprehensive understanding of the phytoplankton size structures found in different water environments. These findings provide a basis for related studies on phytoplankton, including those on the characteristics of the pigment package effect, the primary production proportion of varying size structures, and the response of phytoplankton to climate change.

Acknowledgements

This research was jointly supported by the National Natural Science Foundation of China (No. 41876203, 41576172), the Jiangsu Six Talent Summit Project (No. JY-084), the Qing Lan Project, Natural Science Foundation of Jiangsu Province (SBK2021043413), Natural Science Foundation of the Jiangsu Higher Education Institutions of China (20KJB170028), and Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant no. KYCX21_0975). We show great gratitude to SeaBASS and CSIRO (marlin metadata system) for providing the global cruise data sets and thank NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group for VIIRS *Chl_ocx* Data (accessed on 2021/06/05). The data archiving for the marginal seas of China in this study is underway, and we plan to use a general repository. The in situ pigment data sets are available online (https://seabass.gsfc.nasa.gov/cruise/; https://marlin.csiro.au/), and the VIIRS satellite product is available online (https://oceandata.sci.gsfc.nasa.gov/VIIRS-JPSS1/Mapped/Annual/9km/chl_ocx/).

References

Amante, C., & Eakins, B. W. (2009). *ETOPO1 1 ARC-MINUTE GLOBAL RELIEF MODEL: PROCEDURES, DATA SOURCES AND ANALYSIS*. Paper presented at the NOAA Technical Memorandum NESDIS NGDC-24, National Geophysical Data Center.

Balch, W. M., Gordon, H. R., Bowler, B. C., Drapeau, D. T., & Booth, E. S. (2005). Calcium carbonate measurements in the surface global ocean based on Moderate-Resolution Imaging Spectroradiometer data. *Journal of Geophysical Research: Oceans, 110*(C7). doi:10.1029/2004JC002560

Balch, W. M., Bates, N. R., Lam, P. J., Twining, B. S., Rosengard, S. Z., Bowler, B. C., . . . Rauschenberg, S. (2016). Factors regulating the Great Calcite Belt in the Southern Ocean and its biogeochemical significance. *Global Biogeochemical Cycles, 30*(8), 1124-1144. doi:10.1002/2016gb005414

Behrenfeld, M. J., Randerson, J. T., McClain, C. R., Feldman, G. C., Los, S. O., Tucker, C. J., . . . Pollack, N. H. (2001). Biospheric primary production

during an ENSO transition. *Science, 291*, 2594–2597.

Bracher, & Astrid. (2014). *Phytoplankton pigments measured on water bottle samples during SONNE cruise SO218*.

Brewin, R. J. W., Sathyendranath, S., Hirata, T., Lavender, S. J., Barciela, R. M., & Hardman-Mountford, N. J. (2010). A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecological Modelling, 221*(11), 1472-1483. doi:10.1016/j.ecolmodel.2010.02.014

Brewin, R. J. W., Devred, E., Lavender, S. J., Sathyendranath, S., & Hardman-Mountford, N. J. (2011). Model of phytoplankton absorption based on three size classes. *APPLIED OPTICS, 50*(22), 4535-4549.

Brewin, R. J. W., Sathyendranath, S., Lange, P. K., & Tilstone, G. (2014). Comparison of two methods to derive the size-structure of natural populations of phytoplankton. *Deep Sea Research Part I: Oceanographic Research Papers, 85*, 72-79. doi:10.1016/j.dsr.2013.11.007

Brewin, R. J. W., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Airs, R., & Lamont, T. (2015). Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. *Remote Sensing of Environment, 168*, 437-450. doi:10.1016/j.rse.2015.07.004

Brewin, R. J. W., Ciavatta, S., Sathyendranath, S., Jackson, T., Tilstone, G., Curran, K., . . . Raitsos, D. E. (2017). Uncertainty in Ocean-Color Estimates of Chlorophyll for Phytoplankton Groups. *Frontiers in Marine Science, 4*. doi:10.3389/fmars.2017.00104

Bricaud, A., & Morel, A. (1987). Atmopsheric corrections and interpretation of marine radiances in CZCS imagery: use of reflectance model. *Oceanologica Acta*, 33-50.

Bricaud, A., Claustre, H., Ras, J., & Oubelkheir, K. (2004). Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *Journal of Geophysical Research, 109*(C11). doi:10.1029/2004jc002419

Brotas, V., Brewin, R. J. W., Sá, C., Brito, A. C., Silva, A., Mendes, C. R., . . . Sathyendranath, S. (2013). Deriving phytoplankton size classes from satellite data: Validation along a trophic gradient in the eastern Atlantic Ocean. *Remote Sensing of Environment*.

Devred, E., Sathyendranath, S., Stuart, V., Maass, H., Ulloa, O., & Platt, T. (2006). A two-component model of phytoplankton absorption in the open ocean: Theory and applications. *Journal of Geophysical Research, 111*(C3). doi:10.1029/2005jc002880

Devred, E., Sathyendranath, S., Stuart, V., & Platt, T. (2011). A three component classification of phytoplankton absorption spectra: Application to ocean-color data. *Remote Sensing of Environment, 115*(9), 2255-2266. doi:10.1016/j.rse.2011.04.025

Eleveld, M. A., Pasterkamp, R., Woerd, H. J. v. d., & Pietrzak, J. D. (2008). Remotely sensed seasonality in the spatial distribution of sea-surface suspended particulate matter in the southern North Sea. *Estuarine, Coastal and Shelf Science, 80*, 103–113. doi:10.1016/j.ecss.2008.07.015

Gong, G.-C., Wen, Y.-H., Wang, B.-W., & Liu, G.-J. (2003). Seasonal variation of chlorophyll a concentration, primary production and environmental conditions in the subtropical East China Sea. *Deep Sea Research Part II: Topical Studies in Oceanography, 50*(6-7), 1219-1236. doi:10.1016/s0967-0645(03)00019-5

Hilligsøe, K. M., Richardson, K., Bendtsen, J., S?Rensen, L. L., Nielsen, T. G., & Lyngsgaard, M. M. (2011). Linking phytoplankton community size composition with temperature, plankton food web structure and sea–air CO2 flux. *Deep-Sea Research Part I, 58*(8), 826-838.

Hirawake, T., Takao, S., Horimoto, N., Ishimaru, T., Yamaguchi, Y., & Fukuchi, M. (2011). A phytoplankton absorption-based primary productivity model for remote sensing in the Southern Ocean. *Polar Biology, 34*(2), 291-302. doi:10.1007/s00300-010-0949-y

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research: Oceans, 117*(C1). doi:10.1029/2011jc007395

Huan, Y., Deyong, S., Shengqiang, W., Muhammad, B., Hailong, Z., Zhongfeng, Q., & Yijun, H. (2021). Phytoplankton "Missing" Absorption in Marine Waters: A Novel Pigment Compensation Model for the Packaging Effect. *Journal of Geophysical Research: Oceans*. doi:10.1029/2020jc016458

IOCCG. (2014). *Phytoplankton Functional Types from Space*. Retrieved from Dartmouth, Canada.:

Karageorgis, A. P., & Anagnostou, C. L. (2003). Seasonal variation in the distribution of suspended particulate matter in the northwest Aegean Sea. *Journal of Geophysical Research, 108*(C8), 3274-3304. doi:10.1029/2002JC001672

Kramer, S. J., & Siegel, D. A. (2019). How Can Phytoplankton Pigments Be Best Used to Characterize Surface Ocean Phytoplankton Groups for Ocean Color Remote Sensing Algorithms? *J Geophys Res Oceans, 124*(11), 7557-7574. doi:10.1029/2019JC015604

Kravchishina, M., Lein, A., Burenkov, V., Artemev, V., & Novigatsky, A. (2013). *Distribution and sources of suspended particulate matter in the Kara Sea*. Paper presented at the Complex interfaces under change: Sea-river-groundwater-lake, Gothenburg(SE).

Kuenen, P. (1950). Marine geology. Wiley and sons inc. *New York*.

Lee, Z., Marra, J., Perry, M. J., & Kahru, M. (2015). Estimating oceanic primary productivity from ocean color remote sensing: A strategic assessment. *Journal of Marine Systems*.

Lin, J., Cao, W., Wang, G., & Hu, S. (2014). Satellite-observed variability of phytoplankton size classes associated with a cold eddy in the South China Sea. *Mar Pollut Bull, 83*(1), 190-197. doi:10.1016/j.marpolbul.2014.03.052

Liu, H., Liu, X., Xiao, W., Laws, E. A., & Huang, B. (2021). Spatial and temporal variations of satellite-derived phytoplankton size classes using a three-component model bridged with temperature in Marginal Seas of the Western Pacific Ocean. *Progress in Oceanography, 191*, 102511-102529. doi:10.1016/j.pocean.2021.102511

Liu, S., Qiao, L., Li, G., Li, J., Wang, N., & Yang, J. (2015). Distribution and cross-front transport of suspended particulate matter over the inner shelf of the East China Sea. *Continental Shelf Research, 107*, 92–102. doi:10.1016/j.csr.2015.07.013

Morel, A. (1988). Optical modeling of the upper ocean in relation to its biogenous matter content (case I waters). *Journal of Geophysical Research: Oceans, 93*(C9), 10749-10768. doi:10.1029/jc093ic09p10749

Moutin, T., & Claustre, H. (2013). *Pigments measured on water bottle samples during L'Atalante cruise OLIPAC*. Retrieved from: https://doi.org/10.1594/PANGAEA.804990

Mouw, C. B., & Yoder, J. A. (2010). Optical determination of phytoplankton size composition from global SeaWiFS imagery. *Journal of Geophysical Research, 115*(C12). doi:10.1029/2010jc006337

O'Reilly, J. E., Maritorena, S., O'Brien, M. C., Siegel, D. A., Toole, D., Menzies, D., . . . Culver, M. (2000). *SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3*. Retrieved from

Robert, J., W., Brewin, Giorgio, Dall'Olmo, Shubha, . . . Hardman-Mountford. (2012). Particle backscattering as a function of chlorophyll and phytoplankton size structure in the open-ocean. *Optics Express, 20*(16).

Rosengard, S. Z., Lam, P. J., Balch, W. M., Auro, M. E., Pike, S., Drapeau, D., & Bowler, B. (2015). Carbon export and transfer to depth across the Southern Ocean Great Calcite Belt. *Biogeosciences, 12*, 3953-3971. doi:10.5194/bgd-12-2843-2015

Roy, S., Sathyendranath, S., Bouman, H., & Platt, T. (2013). The global distribution of phytoplankton size spectrum and size classes from their light-absorption spectra derived from satellite data. *Remote Sensing of Environment, 139*, 185-197. doi:10.1016/j.rse.2013.08.004

Sathyendranath, S., Cota, G., Stuart, V., Maass, H., & Platt, T. (2001). Remote sensing of phytoplankton pigments: A comparison of empirical and theoretical approaches. *International Journal of Remote Sensing, 22*(2-3), 249-273. doi:10.1080/014311601449925

Shepard, F. P. (1973). Submarine geology. *journal of geology*.

Sieburth, J. M., Smetacek, V., & Lenz, J. (1978). Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions 1. *Limnology and Oceanography, 23*(6), 1256-1263. doi:10.4319/lo.1978.23.6.1256

Smith, H. E. K., Poulton, A. J., Garley, R., Hopkins, J., Lubelczyk, L. C., Drapeau, D. T., . . .

Balch, W. M. (2017). The influence of environmental variability on the biogeography of coccolithophores and diatoms in the Great Calcite Belt. *Biogeosciences, 14*(21), 4905-4925. doi:10.5194/bg-14-4905-2017

Sun, D., Huan, Y., Qiu, Z., Hu, C., Wang, S., & He, Y. (2017). Remote-Sensing Estimation of Phytoplankton Size Classes From GOCI Satellite Measurements in Bohai Sea and Yellow Sea. *Journal of Geophysical Research: Oceans, 122*(10), 8309-8325. doi:10.1002/2017jc013099

Sun, D., Huan, Y., Wang, S., Qiu, Z., Ling, Z., Mao, Z., & He, Y. (2019). Remote sensing of spatial and temporal patterns of phytoplankton assemblages in the Bohai Sea, Yellow Sea, and east China sea. *Water Res, 157*, 119-133. doi:10.1016/j.watres.2019.03.081

Sun, X., Shen, F., Liu, D., Bellerby, R. G. J., Liu, Y., & Tang, R. (2018). In Situ and Satellite Observations of Phytoplankton Size Classes in the Entire Continental Shelf Sea, China. *Journal of Geophysical Research: Oceans, 123*(5), 3523-3544. doi:10.1029/2017jc013651

Svendsen, S., & Shrum, C. (1995). An estimate of global primary production in the ocean from satellite radiometer data. *Journal of Plankton Research, 17*(6).

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research, 111*(C8). doi:10.1029/2005jc003207

Uitz, J., Huot, Y., Bruyant, F., Babin, M., & Claustre, H. (2008). Relating phytoplankton photophysiological properties to community structure on large scales. *Limnology and Oceanography, 53*(2), 614-630. doi:10.4319/lo.2008.53.2.0614

Uitz, J., Stramski, D., Reynolds, R. A., & Dubranna, J. (2015). Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. *Remote Sensing of Environment, 171*, 58-74. doi:10.1016/j.rse.2015.09.027

Vidussi, F., laustre, C., Manca, B. B., Luchetta, A., & Marty, J.-C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea. *Journal of Geophysical Research, 106*(C9), 19939-19956. doi:10.1029/1999JC000308

Wang, G., Cao, W., Wang, G., & Zhou, W. (2013). Phytoplankton size class derived from phytoplankton absorption and chlorophyll-a concentrations in the northern South China Sea. *Chinese Journal of Oceanology and Limnology, 31*(4), 750-761. doi:10.1007/s00343-013-2291-z

Wang, S., Ishizaka, J., Hirawake, T., Watanabe, Y., Zhu, Y., Hayashi, M., & Yoo, S. (2015). Remote estimation of phytoplankton size fractions using the spectral shape of light absorption. *Opt Express, 23*(8), 10301-10318. doi:10.1364/OE.23.010301

Werdell, P. J., & Bailey, S. W. (2005). An improved in-situ bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment, 98*(1), 122-140. doi:10.1016/j.rse.2005.07.001

Werdell, P. J., Boss, E., Franz, B. A., Brando, V. E., Bailey, S. W., Lavender, S. J., . . . Mangin, A. (2013). Generalized ocean color inversion model for retrieving marine inherent optical properties. *APPLIED OPTICS, 52*(10), 2019-2037.

Xu, M., & Chen, Y. (1999). *Geological Oceanography*. Xiamen, China.: Xiamen University Press.

Yan, B., Tingwei, C., Lian, F., Chengfeng, L., Zongping, L., Xiaoju, P., . . . Yunlin, Z. (2019). *Introduction to oceancolor* (z. p. Lee Ed.). Xiamen, China.: Xiamen University Press.

Zhang, T., & Shi, Y. (2005). A Method to Classify Case I and Case II Waters. *Periodical of Ocean University of China, 35*(5), 849-853.