

**Key Points:**

- An algorithm for globally retrieving multiple phytoplankton functional types (PFTs) is retuned by including sea surface temperature data
- Improved satellite PFT products are provided using the GlobColour merged ocean color products
- The per-pixel uncertainty of the satellite PFT products is assessed in detail by accounting for errors from input data and model parameters

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

H. Xi,  
[Hongyan.Xi@awi.de](mailto:Hongyan.Xi@awi.de)

**Citation:**

Xi, H., Losa, S. N., Mangin, A., Garnesson, P., Bretagnon, M., Demaria, J., et al. (2021). Global chlorophyll *a* concentrations of phytoplankton functional types with detailed uncertainty assessment using multisensor ocean color and sea surface temperature satellite products. *Journal of Geophysical Research: Oceans*, 126, e2020JC017127. <https://doi.org/10.1029/2020JC017127>

Received 23 DEC 2020  
 Accepted 10 APR 2021

© 2021. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](https://creativecommons.org/licenses/by-nc/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## Global Chlorophyll *a* Concentrations of Phytoplankton Functional Types With Detailed Uncertainty Assessment Using Multisensor Ocean Color and Sea Surface Temperature Satellite Products

Hongyan Xi<sup>1</sup> , Svetlana N. Losa<sup>1,2</sup>, Antoine Mangin<sup>3</sup>, Philippe Garnesson<sup>3</sup>, Marine Bretagnon<sup>3</sup>, Julien Demaria<sup>3</sup>, Mariana A. Soppa<sup>1</sup>, Odile Hembise Fanton d'Andon<sup>3</sup>, and Astrid Bracher<sup>1,4</sup> 

<sup>1</sup>Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany, <sup>2</sup>Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia, <sup>3</sup>ACRI-ST, Sophia Antipolis Cedex, France, <sup>4</sup>Institute of Environmental Physics, University of Bremen, Bremen, Germany

**Abstract** First, we retune an algorithm based on empirical orthogonal functions (EOFs) for globally retrieving the chlorophyll *a* concentration (Chl-*a*) of phytoplankton functional types (PFTs) from multisensor merged ocean color (OC) products. The retuned algorithm, referred to as EOF-SST hybrid algorithm, is improved by: (i) using 23% more matchups between the updated global in situ pigment database and satellite remote sensing reflectance ( $R_{rs}$ ) products, and (ii) including sea surface temperature (SST) as an additional input parameter. In addition to the Chl-*a* of the six PFTs (diatoms, haptophytes, dinoflagellates, green algae, prokaryotes, and *Prochlorococcus*), the fractions of prokaryote and *Prochlorococcus* Chl-*a* to total Chl-*a* (TChl-*a*), are also retrieved by the EOF-SST hybrid algorithm. Matchup data are separated for low and high-temperature regimes based on different PFT dependences on SST, to establish SST-separated hybrid algorithms which demonstrate further improvements in performance as compared to the EOF-SST hybrid algorithm. The per-pixel uncertainty of the retrieved TChl-*a* and PFT products is estimated by taking into account the uncertainties from both input data and model parameters through Monte Carlo simulations and analytical error propagation. The algorithm and its method to determine uncertainties can be transferred to similar OC products until today, enabling long-term continuous satellite observations of global PFT products. Satellite PFT uncertainty is essential to evaluate and improve coupled ecosystem-ocean models which simulate PFTs, and furthermore can be used to directly improve these models via data assimilation.

**Plain Language Summary** Phytoplankton in the sunlit layer of the ocean contribute approximately 50% to global primary production. They act as the base of the marine food web fueling fisheries, and also regulate key biogeochemical processes such as exporting carbon to the deep ocean. Phytoplankton contain various taxonomic groups that function differently in the marine ecosystem. The global phytoplankton can be observed from space by analyzing the signal leaving from the water surface recorded by the ocean color (OC) sensors onboard the satellites. Based on an updated large global data set, satellite data from different OC sensors, and sea surface temperature data, we adapted our previous approach to improve the quantification of the biomass of the main six phytoplankton groups on a global scale. The uncertainty of the satellite products of the phytoplankton groups is calculated by considering the errors propagated from the satellite data and the model parameters. This approach for quantifying different phytoplankton groups, together with the uncertainty assessment, can be extended to other similar OC satellite data which cover different time periods, to ultimately generate long-term global distribution maps of multiple phytoplankton groups. This information will help the modelers to predict better the phytoplankton changes in the future.

### 1. Introduction

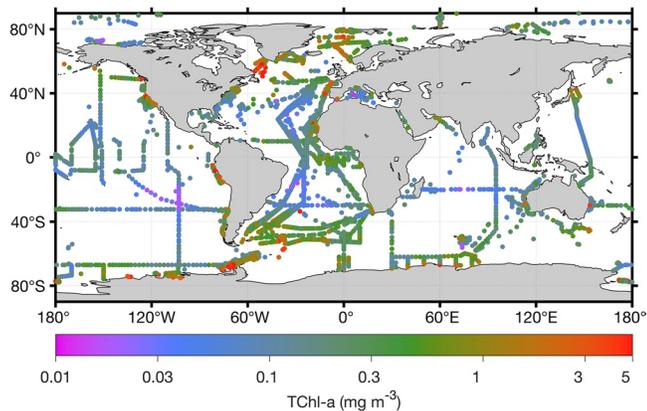
As phytoplankton play a fundamental role in marine food webs and biogeochemical cycling, their community structure and taxonomic composition have been widely investigated in recent decades through various observational methods and ecological modeling (e.g., Falkowski et al., 2003; IOCCG, 2014; Le Qué

et al., 2005). With a vast amount of quality-controlled ocean color (OC) remote sensing data now available, methods to infer the dominance or fraction of phytoplankton groups, size classes (PSCs) and phytoplankton functional types (PFTs) on a large scale both in space and time have evolved. Retrieval algorithms for phytoplankton composition have generally been developed based on both in situ measurements and satellite products, the former providing the ground truth information at specific time and location, but with the latter regularly repeated, high spatial resolution observations can be achieved.

Radiance or reflectance signals leaving the ocean surface and measured by a satellite radiometer contain phytoplankton pigment information that can be related to community structure and size classes (Bracher et al., 2017; Mouw et al., 2017). Therefore, spectral-based approaches have been developed to retrieve the concentrations of phytoplankton chlorophyll, pigments, and multiple PFTs from space (e.g., Alvain et al., 2005, 2008; Bracher et al., 2009; Correa-Ramirez et al., 2018; Lange et al., 2020; Werdell et al., 2014; Xi et al., 2020). One such approach that has proven efficient is based on the empirical orthogonal function (EOF) analysis on the spectral  $R_{rs}$  or water-leaving radiance. By reducing the high dimensionality of the spectral data, the dominant signals that best describe the variance of the structures lying in the spectra can be assessed to establish statistical models for predicting OC metrics, including various phytoplankton pigment and PFT chlorophyll *a* concentrations (Chl-*a*) (e.g., Bracher et al., 2015; Craig et al., 2012; Lange et al., 2020; Lubac & Loisel, 2007; Soja-Woźniak et al., 2017; Taylor et al., 2013; Xi et al., 2020). Approaches based on EOF analysis also exhibited equivalent skill with a little downgrading of the performance when applied to reduced spectral resolution (Bracher et al., 2015), enabling their wide applicability to previous (e.g., SeaWiFS and MERIS) and current (e.g., MODIS, VIIRS, and OLCI) multispectral OC sensors and their merged products. In addition, as these approaches are usually trained to retrieve OC metrics and PFT information directly from the satellite spectral data, no prior knowledge of the phytoplankton biomass or inherent optical properties (IOPs) is required in its application. This makes the implementation of such approaches straightforward and practical for satellite OC products.

The EOF-based approach proposed by Xi et al. (2020) has been shown to provide reliable Chl-*a* retrievals of multiple PFTs on the global scale, through intercomparisons with other satellite-derived PFT and PSC products. However, PFT retrievals by Xi et al. (2020) showed rather low performance for prokaryotic phytoplankton. Incorporation of additional environmental parameters which are globally available from satellite measurements, such as optical depth, sea surface temperature (SST), wind stress, and light availability, have the potential to improve PSC prediction accuracy. For instance, Brewin et al. (2015) investigated the influence of light in the mixed layer on the parameters of the three-component PSC model (abundance-based model) of Brewin et al. (2010), and modified the model to better describe the relationship between phytoplankton size structure and total chlorophyll with varying light conditions. Ward (2015) and Brewin, Ciavatta, et al. (2017) both incorporated temperature dependence into the three-component model and improved the model's ability in representing Chl-*a* concentrations in all three PSCs using satellite estimates of SST and total Chl-*a* concentration. Moore and Brown (2020) assessed the impact of separately including SST and estimates of light level in the surface mixed-layer on the skill of abundance-based PSC models. Lange et al. (2020) also considered SST in their spectral-based algorithm using principal component regressions to estimate the surface abundance of picophytoplankton across the Atlantic Ocean.

There has been an emerging trend of the combined use of in situ data, satellite observations, ecosystem modeling (Losa et al., 2019), as well as PFT or PSC data assimilation (Pradhan et al., 2020; Xiao & Friedrichs, 2014), to allow comprehensive monitoring and predictions of phytoplankton community structure. Satellite-derived phytoplankton group-specific products are also expected to be useful for validation of ecosystem model results (e.g., de Mora et al., 2016; Dutkiewicz et al., 2015; Hirata et al., 2013; Holt et al., 2014; Pradhan et al., 2019; Ward et al., 2012). One major challenge of fulfilling these tasks is associating the uncertainty with the satellite-derived PFT products (Bracher et al., 2017). Uncertainty estimates have been well formulated and generated for other common OC algorithms that use satellite radiance and reflectance data to derive OC products such as Chl-*a* concentration, diffuse attenuation coefficient, and IOPs (McKinna et al., 2019; Werdell et al., 2018). Though various approaches have been proposed to globally estimate satellite phytoplankton group products (Mouw et al., 2017), only the study by Brewin, Ciavatta, et al. (2017) has provided estimates of uncertainty on a per-pixel basis for the North Atlantic Ocean. Uncertainty assessment can be carried out via two methods: validation through comparison of the satellite retrievals with in



**Figure 1.** Distribution of total chlorophyll *a* concentration (TChl-*a*, the sum of monovinyl chlorophyll *a*, divinyl chlorophyll *a*, chlorophyll *a* allomers, chlorophyll *a* epimers, and chlorophyllide *a*) from the quality-controlled in situ pigment database (2002–2012) used in this study ( $n = 8,840$ ).

situ data (e.g., Antoine et al., 2008; Sathyendranath et al., 2019), or error propagation by accounting for the uncertainties in the inputs and model parameters. Due to the sparse distribution of in situ measurements that restricts validation for the uncertainty estimation (Mélin & Franz, 2014), error propagation analysis has now been widely used not only to understand the sensitivity of model inputs and parameters to model outputs, but also to estimate pixel-by-pixel uncertainty (e.g., Brewin, Tilstone, et al., 2017; Kostadinov et al., 2016; Lee et al., 2010; Maritorena et al., 2010; Qi et al., 2017).

In this study, we improve the previously developed EOF-based algorithm of Xi et al. (2020) for global retrievals of multiple PFT quantities by (1) including more matchup data between the in situ pigment data set and satellite  $R_{rs}$  data from merged OC products, (2) accounting for the SST in the retrieval scheme, and (3) investigating the influence of SST on the model parameters and retrieved PFTs with the goal of establishing a set of EOF-SST hybrid algorithms to improve the retrievals of TChl-*a*, Chl-*a* of six PFTs and the fractions of two prokaryotic phytoplankton. By applying the hybrid algorithms to the merged OC products, we generate global distribution maps of the satellite-retrieved PFT quantities, and present a method to derive the per-pixel uncertainty propagated from both the

inputs and retrieval model for each PFT quantity by combining Monte Carlo (MC) simulations and an analytical approach.

## 2. Data and Methods

### 2.1. Data Sets

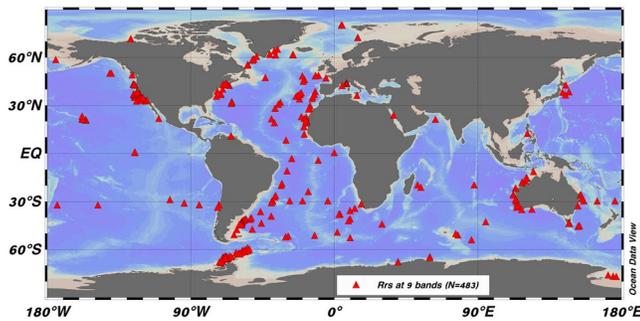
#### 2.1.1. In Situ Data Set of Phytoplankton Pigments and Data Screening

We updated the large global and open ocean (water depth >200 m) phytoplankton pigment data set (spanning 2002–2012) from Losa et al. (2017) used in Xi et al. (2020) analyzed by High-Performance Liquid Chromatography (HPLC), by adding recently published HPLC pigment data (as of February 2020) from SeaBASS, PANGAEA, British Oceanographic Data Centre (BODC), and Open Access to Ocean Data (AODN) from Australia (data sources in Table S1). All collected data were quality controlled following the method by Aiken et al. (2009); only pigment concentrations greater than the HPLC detection limit of  $0.001 \text{ mg m}^{-3}$  were included. In the database, all required pigments for the PFT Chl-*a* calculation were used (fucoxanthin, peridinin, 19′hexanoyloxy-fucoxanthin, 19′butanoyloxy-fucoxanthin, alloxanthin, total chlorophyll *b*, zeaxanthin, and divinyl chlorophyll *a*, reference to Section 2.1.3.1). While investigating the pigment data set, we found that data collected from the Palmer Long-Term Ecological Research (LTER, <https://pal.lternet.edu/>) in the west of the Antarctic Peninsula showed very high zeaxanthin concentrations (up to  $2.7 \text{ mg m}^{-3}$ ) for multiple years during 2002–2012 compared to data from the other cold waters in the Southern Ocean. We considered these distinct zeaxanthin measurements from Palmer LTER as nonrepresentative in the Southern Ocean and screened the pigment samples which had zeaxanthin concentration higher than the mean ( $0.013 \text{ mg m}^{-3}$ ) of zeaxanthin concentration in the other waters in the Southern Ocean. Details of the data screening and supporting references and communications are provided in the supporting information (Text S1 and Figures S1–S3). A total of 8,840 sets of pigment data were finally used in this study and their distribution is shown in Figure 1.

#### 2.1.2. Satellite Data

##### 2.1.2.1. Satellite Ocean Color Products

GlobColour (<http://www.globcolour.info/>) has provided various OC products from different sensors, including SeaWiFS, MODIS-Aqua, MERIS, VIIRS-NPP, and Sentinel-3A OLCI. In this study, we used the SeaWiFS-MODIS-MERIS merged normalized remote sensing reflectance ( $R_{rs}$ ) Level-3 (L3) product (hereafter referred to as merged product) which covers the period from July 2002 to April 2012 from the GlobColour



**Figure 2.** Geographical locations of single-pixel matchups of GlobColour merged  $R_{rs}$  at nine bands with in situ PFT and satellite SST data. PFT, phytoplankton functional type; SST, sea surface temperature.

data archive (more details in ACRI-ST GlobColour Team et al. [2017]). As in Xi et al. (2020), the daily merged product with 4-km resolution was used for matchup extraction and the monthly merged product with 25-km resolution was used for algorithm application. Since this study focuses mainly on oceanic waters, shelf and coastal waters (<200 m) were masked out in the OC products following Hirata et al. (2011) using the ETOPO1 bathymetry (Amante & Eakins, 2009).

### 2.1.2.2. Sea Surface Temperature Data

The SST product used in this study was CMEMS OSTIA (Operational SST and Ice Analysis) reprocessed analysis product, which is quality-controlled and available on the CMEMS (Copernicus Marine Environment Monitoring Service, <https://marine.copernicus.eu/>) platform, referenced as SST\_GLO\_SST\_L4\_REP\_OBSERVATIONS\_010\_011. The CMEMS OSTIA reprocessed analysis product is an interpolated product based on in situ measurements and satellite observations from both infra-red and

microwave data on a global regular grid at 0.05° resolution (Donlon et al., 2012; Worsfold et al., 2020). The daily SST product from July 2002 to April 2012 was acquired and gridded to 4-km resolution. As for  $R_{rs}$ , monthly mean SST from 2002 to 2012 with 25-km resolution were also processed as input for deriving global satellite PFT products.

### 2.1.3. Input Data for PFT Retrieval Algorithm

#### 2.1.3.1. In Situ PFT Chl-a and Fraction Derived From Diagnostic Pigment Analysis

As described in Xi et al. (2020), Chl-a of PFTs based on in situ pigment data were derived using an updated diagnostic pigment analysis (DPA) method (Losa et al., 2017; Soppa et al., 2014), that was originally developed by Vidussi et al. (2001), adapted in Uitz et al. (2006) and further refined by Hirata et al. (2011) & Brewin et al. (2015). We used pigment concentrations from the in situ database mentioned in Section 2.1.1 to derive the Chl-a of six PFTs—diatoms, dinoflagellates, haptophytes, green algae, prokaryotes, and *Prochlorococcus*, with fucoxanthin, peridinin, 19'hexanoyloxy-fucoxanthin, total chlorophyll *b*, zeaxanthin and divinyl chlorophyll *a* as their diagnostic pigment (DP), respectively. The partial coefficients of the DPs used in the updated DPA method were derived using a large global pigment data set as detailed in Table S1 in Losa et al. (2017), which were shown to be in good agreement with previous studies. We used these coefficients rather than determining new ones because the majority (~80%) of the in situ pigment data used in this study are a subset of the global data from 1988 to 2012 compiled in Losa et al. (2017). Due to the weak retrieval performance of prokaryote and *Prochlorococcus* Chl-a in Xi et al. (2020), in this study, we included the fractions of prokaryotes (*f*-prokaryotes) and *Prochlorococcus* (*f*-*Prochlorococcus*) to TChl-a, attempting to get improved retrievals of these two PFTs. PFT Chl-a lower than 0.005 mg m<sup>-3</sup> were excluded due to high uncertainty (Xi et al., 2020) and the corresponding fractions of prokaryotes and *Prochlorococcus* were also excluded.

#### 2.1.3.2. Matchups Between Satellite SST and In Situ PFT Data

SST matchups were extracted by matching data that were spatially colocalized and temporally (on a daily basis) coincident with the in situ PFT measurements. A 3×3 macro-pixel centered on the in situ measurement was considered. If the standard deviation compared to the mean value within this macro-pixel was lower than 25%, then the macro-pixel was considered suitable for the matchup. The arithmetic median of the macro-pixel was defined as the SST value for the in situ site.

#### 2.1.3.3. Matchups of $R_{rs}$ Merged Product to In Situ PFT and Satellite SST Data

Matchups of satellite  $R_{rs}$  to in situ PFT and satellite SST data were extracted from global 4-km daily merged products. The same extraction, averaging, and quality control protocol as followed in Xi et al. (2020) was used to derive matchups. As described in Xi et al. (2020), due to more matchup points and equivalent retrieval performance compared to 3×3 matchups, single-pixel matchups were used as the input data for the final retrieval approach (Figure 2). Matchup data with the nine spectral bands of 412, 443, 490, 510, 531,

**Table 1**  
List of the Nine Bands From Sensors SeaWiFS, MODIS, and MERIS Used in the GlobColour Merged Products

| Sensors | Center wavebands (nm) |     |     |     |     |     |     |     |     |
|---------|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|
|         | 412                   | 443 | 490 | 510 | 531 | 547 | 555 | 670 | 678 |
| SeaWiFS | ×                     | ×   | ×   | ×   |     |     | ×   | ×   |     |
| MODIS   | ×                     | ×   | ×   |     | ×   | ×   | ×   | ×   | ×   |
| MERIS   | ×                     | ×   | ×   | ×   |     |     | ×   | ×   |     |

*Note.* \*There was no band at 555 nm for MERIS itself, but the GlobColour Team provided for MERIS the 555 nm through an interspectral conversion made by using:  $R_{rs}(555) = R_{rs}(560) * (1.02542 - 0.03757 * y - 0.00171 * y^2 + 0.0035 * y^3 + 0.00057 * y^4)$ , where  $y = \log_{10}(\text{CHL1})$  and CHL1 is the total Chl-a concentration estimated by OC4 (ACRI-ST GlobColour Team et al., 2017). With this conversion,  $R_{rs}$  at 555 nm for MERIS were also included in our study.

547, 555, 670, and 678 nm from the merged products were chosen as the algorithm input data (Table 1). A total of 483 sets of matchup data covering the global ocean were extracted (Figure 2).

## 2.2. Algorithm Retuning

The EOF-based PFT retrieval algorithm development and performance assessment were detailed in Xi et al. (2020). Using the updated matchup data, the Xi et al. (2020) algorithm was retuned and statistically assessed as detailed below. Figure 3 shows the scatterplots of the matchup data of SST to TChl-a, the six PFT Chl-a, and the two PFT fractions. Generally, TChl-a, Chl-a of diatoms, haptophytes, dinoflagellates, and green algae show a decrease with increasing SST. However, Prokaryote and *Prochlorococcus* Chl-a and their fractions show positive correlation with SST. The statistically significant correlations indicate that introducing SST as an additional term into the Xi et al. (2020) algorithm (see Section 2.2.1) might improve the algorithm's performance. With a 10-point running mean applied to the data, there is a noticeable shift in the trends of most

PFT quantities as a function of SST at approximately 13 °C. This further led us to establish for  $SST < 13$  °C a different EOF-SST hybrid algorithm than for  $SST \geq 13$  °C (see Section 2.2.2). *Prochlorococcus* data, as an exception from other PFTs, are rarely recorded in high latitudes with low temperature. Our final matchup data set compiled in Section 2.1.3.3 contained few ( $n = 2$ ) divinyl chlorophyll *a* (a marker pigment of *Prochlorococcus*) measurements with low concentrations ( $0.001 \text{ mg m}^{-3}$ ) when  $SST < 10$  °C, which is in agreement with the observations by Flombaum et al. (2013). Therefore, we excluded the regions where SST is below 10 °C and did not consider a different EOF-SST hybrid algorithm with  $SST < 13$  °C for the Chl-a and fraction retrievals of *Prochlorococcus*.

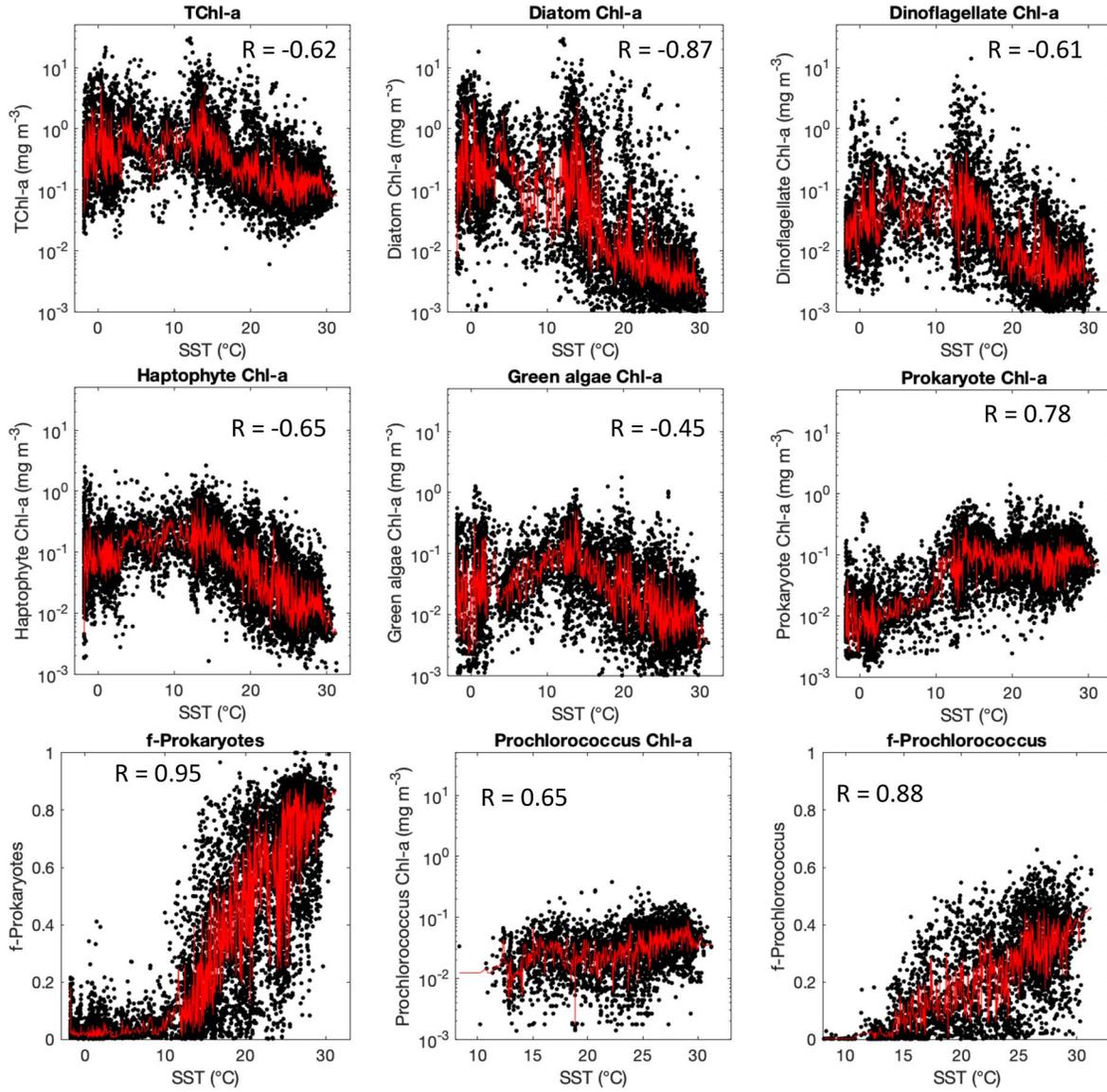
### 2.2.1. Adapted EOF-SST Hybrid PFT Algorithm Based on the Whole Data Set

The input data set used in the EOF-SST hybrid algorithm was the matchup data set that included the collocated nine-band  $R_{rs}$  from the merged products, satellite SST data, and in situ PFT data. Figure 4 depicts the flowchart of the EOF-SST hybrid algorithm, in which the EOF analysis remained unchanged by still using singular value decomposition (SVD) to decompose the (standardized)  $R_{rs}$  spectra into the EOF scores ( $U$ ), singular values ( $\Lambda$ ), and EOF loadings ( $V$ ) as in Xi et al. (2020). Now, when formulating the regression models of PFTs, we introduced SST as an additional term together with the column vectors  $u_{1,2,\dots,n}$  in  $U$ . Similar to Xi et al. (2020), we applied a stepwise routine to obtain the smaller regression model by removing least significant variables in  $U$  through minimization of the Akaike information criterion (AIC). The adapted regression model is expressed as

$$\ln(C_p) = a_0 + a_1 u_1 + a_2 u_2 + \dots + a_n u_n + a_{SST} SST \quad (1)$$

where  $a_0$  is the intercept,  $a_{1,2,\dots,n}$ , and  $a_{SST}$  are the regression coefficients for the selected EOF scores and SST, respectively. With the adapted regression model, the same steps for cross-validation and model assessment of Xi et al. (2020) were carried out to test the robustness of the fitted model: the whole collocated data set was randomly split into two subsets—the first subset containing 80% of the data was used for model fitting/training and the rest 20% was used for prediction. The procedure was run for 500 permutations to (1) record for each permutation the model parameters for further uncertainty assessment, and (2) generate a final statistical assessment based on the statistics of the model performance derived from each permutation.

For the model assessment, we considered the slope ( $S$ ) and the intercept ( $a$ ) of the generalized linear model (GLM) regression and coefficient of determination ( $R^2$ ), which were based on the log-scaled PFT predictions against the log-scaled in situ PFT data. We also considered the root-mean-square difference (RMSD), the median percent difference (MDPD), and the bias that were based on the non-log-transformed concentration data. These metrics were expressed as

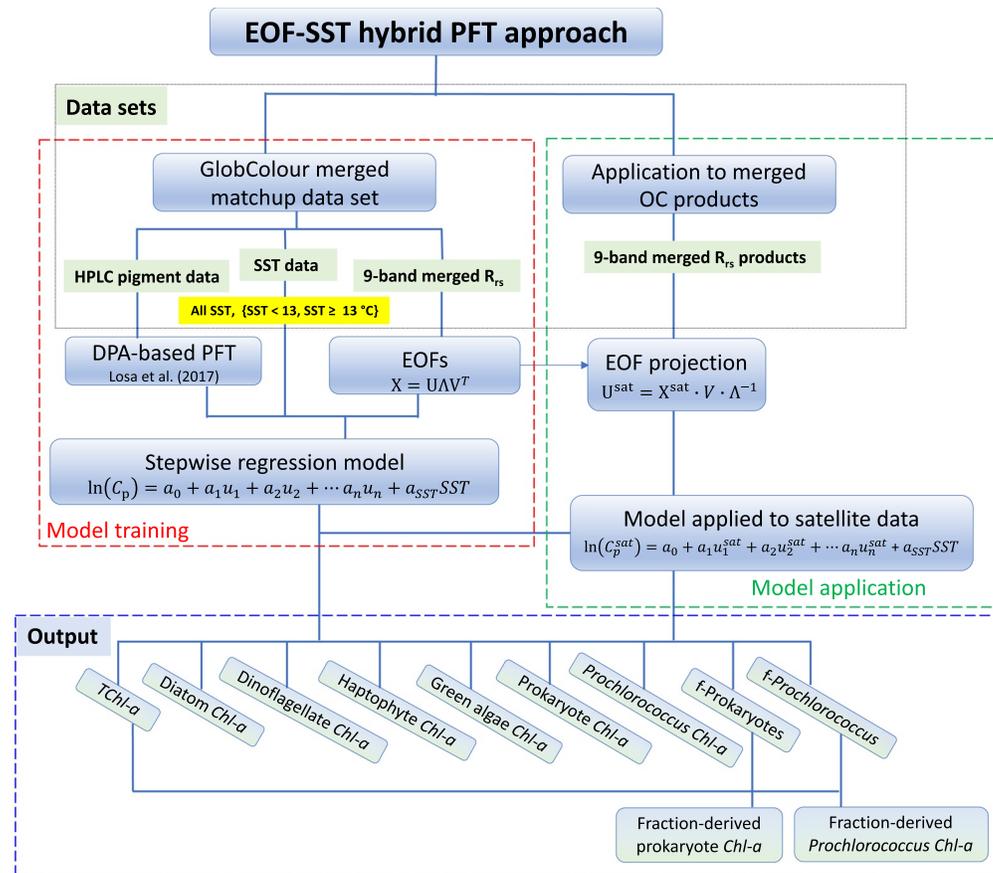


**Figure 3.** Scatterplots of in situ TChl-a, PFT Chl-a, and fractions of prokaryotes and *Prochlorococcus* versus collocated satellite SST data. The correlation coefficient ( $R$ ) was calculated based on the 10-point running mean (red curve). PFT, phytoplankton functional type; SST, sea surface temperature.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (C_{pi} - C_{oi})^2}{N}}, \quad (2)$$

$$\text{MDPD} = \text{Median of } \left[ \left| \frac{C_{pi} - C_{oi}}{C_{oi}} \right| \times 100 \right], i = 1, \dots, N, \quad (3)$$

$$\text{bias} = \frac{100}{N} \sum_{i=1}^N \frac{(C_{pi} - C_{oi})}{C_{oi}}, \quad (4)$$



**Figure 4.** Flowchart illustrating the EOF-SST hybrid algorithm and the SST-separated hybrid algorithms for predicting TChl-a, Chl-a of six PFTs, and two fractions with GlobColour merged product. The red dashed-line box depicts the model training with the pigment-satellite matchup data; the green dashed-line box depicts the model application to satellite products and the blue dashed-line box shows model outputs (i.e., the predicted PFT quantities). EOF, empirical orthogonal function; PFT, phytoplankton functional type; SST, sea surface temperature.

where  $N$  is the number of observations of PFTs ( $C_o$ ) and the corresponding predictions ( $C_p$ ). Meanwhile, the cross-validation statistics ( $R^2_{cv}$ ,  $RMSD_{cv}$ , and  $MDPD_{cv}$ ), representing both the model robustness and compromised model performance, were also determined by taking the mean of the statistical parameters  $R^2$ ,  $RMSD$ , and  $MDPD$  from all permutations, respectively.

### 2.2.2. SST-Separated Hybrid PFT Algorithms

Given the difference in SST-PFT relationships between the two data sets (SST < 13 °C and SST ≥ 13 °C, Figure 3), for all PFT quantities (excluding *Prochlorococcus* Chl-a and *f-Prochlorococcus*), we separated the matchup data set at 13 °C as a divide and established two specific EOF-SST hybrid algorithms using the two data sets (referred to as SST-separated hybrid algorithms) following Section 2.2.1 (Figure 4). Note that the performance of the SST-separated hybrid algorithms was evaluated statistically based on all the predictions and in situ data to be consistent with that for the EOF-SST hybrid algorithm.

### 2.2.3. Application of Algorithms

The established algorithms were applied to the satellite  $R_{rs}$  data from the merged products (Section 2.1.2) to retrieve PFTs globally (Figure 4). By projecting the satellite  $R_{rs}$  data onto the EOF loadings ( $V$ ), a new set of EOF scores ( $U^{sat}$ ) was derived and was then used for the global PFT prediction together with SST as an additional term in the fitted model in Equation 1, where  $a_0$  and  $a_{1,2,\dots,n}$  were obtained in the step of model training.

**Table 2**  
Summary of Data Sets Included in Section 2 for Various Analyses/Usages

| Type of data set                           | Purpose of use                                      | Subsection      | Number | Data involved |                             |     |                               |
|--|---|-----------------|--------|---------------|-----------------------------|-----|-------------------------------|
|  |   |                 |        | HPLC pigments | TChl-a and DPA-derived PFTs | SST | R <sub>rs</sub> at nine bands |
| In situ pigment database                   | Original in situ database                           | Section 2.1.3.1 | 8,840  | ×             |                             |     |                               |
| Collocated SST and PFTs                    | PFT dependence on SST                               | Section 2.1.3.2 | 8,840  |               | ×                           | ×   |                               |
| Matchups of PFTs, SST, and R <sub>rs</sub> | EOF-SST hybrid algorithm                            | Section 2.1.3.3 | 483    |               | ×                           | ×   | ×                             |
| Matchups of PFTs, SST, and R <sub>rs</sub> | SST-separated algorithm for waters with SST ≥ 13 °C | Section 2.2.1   | 368    |               | ×                           | ×   | ×                             |
| Matchups of PFTs, SST, and R <sub>rs</sub> | SST-separated algorithm for waters with SST < 13 °C | Section 2.2.2   | 115    |               | ×                           | ×   | ×                             |
| Training data set from matchups            | Monte Carlo simulations                             | Section 2.3.2   | 242    |               | ×                           | ×   | ×                             |
| Test data set from matchups                | Uncertainty validation                              | Section 2.3.3   | 241    |               | ×                           | ×   | ×                             |

As various categories of data sets were compiled in Section 2 for different analyses regarding PFT global retrieval algorithms and uncertainty assessment, Table 2 summarizes the details of the data sets compiled for this study. It should be noted that the numbers of the data points indicated in Table 2 are based on valid TChl-a data. This may be lower for specific PFT data sets (indicated in Figures 5 and 6 and Tables 4 and 5 in Section 3.1) as PFTs can be absent in some samples.

### 2.3. Uncertainty Assessment of PFT Retrieval

To quantify the uncertainty of the satellite PFT retrievals, we considered the uncertainties propagated from the input satellite R<sub>rs</sub> ( $\sigma_{Rrs}$ ) and SST ( $\sigma_{SST}$ ), and the uncertainty of the model/algorithm parameters ( $\sigma_a$ ). Other uncertainty sources exist, that is, in situ HPLC pigment measurement error and uncertainties in the DPA-derived PFT data resulting from incorrect assignment of PFTs from marker pigments. In the current study, we could not quantify the combined uncertainty from HPLC measurements and the DPA-derived PFTs due to limited information and therefore did not include this error source in the uncertainty assessment, although this error source should still be kept in mind even though they could not be quantified in this study. Detailed discussion regarding this aspect has been elaborated in Section 3.3.4 following the results of PFT uncertainty.

All computations of the uncertainties in this study were based on the logarithmic transformed data following conventional practice in the field of OC (Sathyendranath et al., 2020). However, we used natural logarithms instead of the common (base 10) logarithms, as our algorithm was developed based on the natural logarithms. For comparison with other studies, the common logarithmic uncertainty can be obtained by dividing our uncertainty by  $\ln(10)$ , that is, approximately a factor of 2.3. For brevity only the uncertainty derived based on the whole EOF-SST hybrid algorithm is presented as a general approach to quantify and consolidate the PFT uncertainty from different error sources.

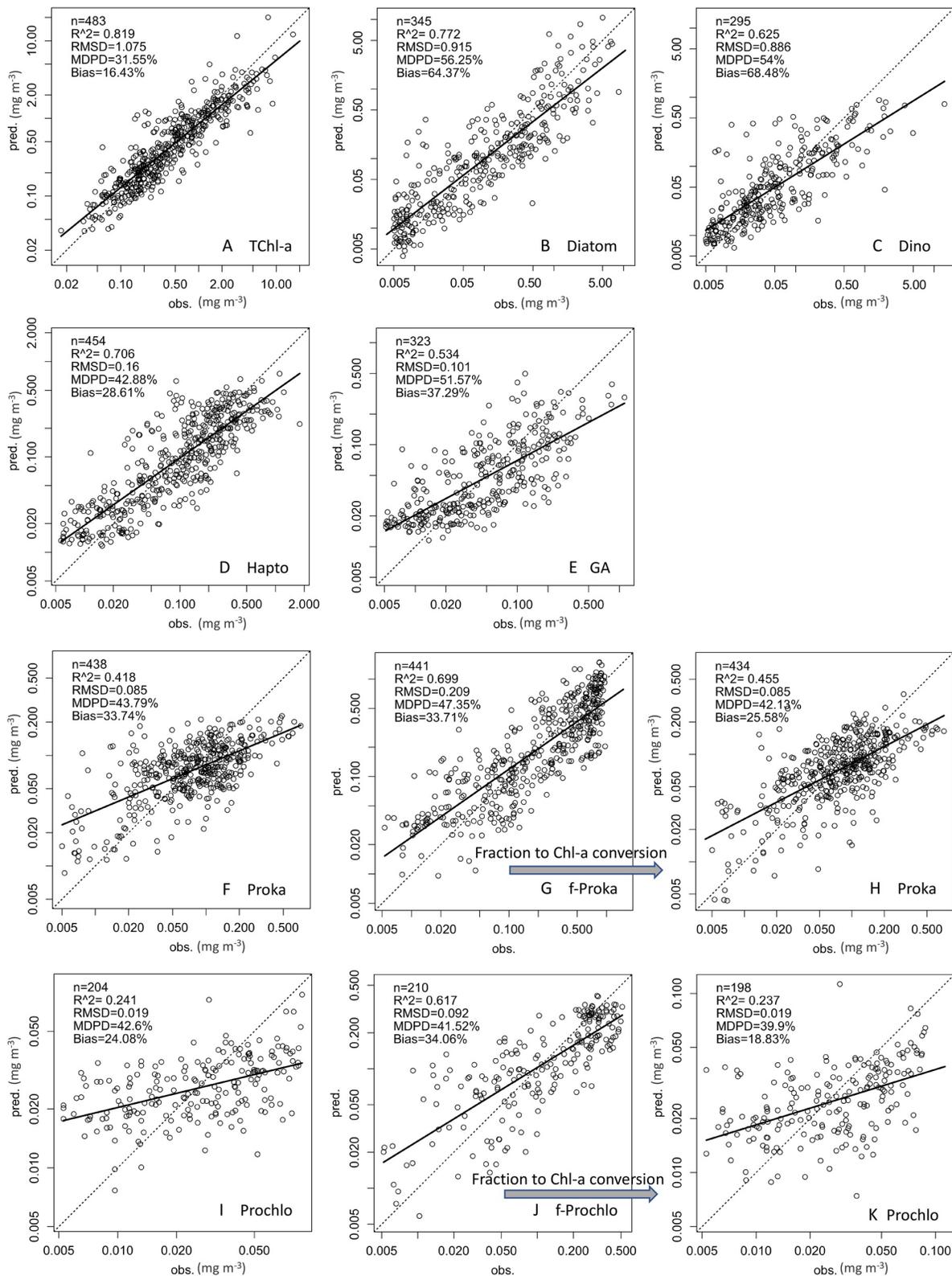
#### 2.3.1. Structure of the Uncertainty Propagation

With the EOF-SST hybrid retrieval models expressed in Equation 2, the retrieval model applied to the satellite data can be written in the following form:

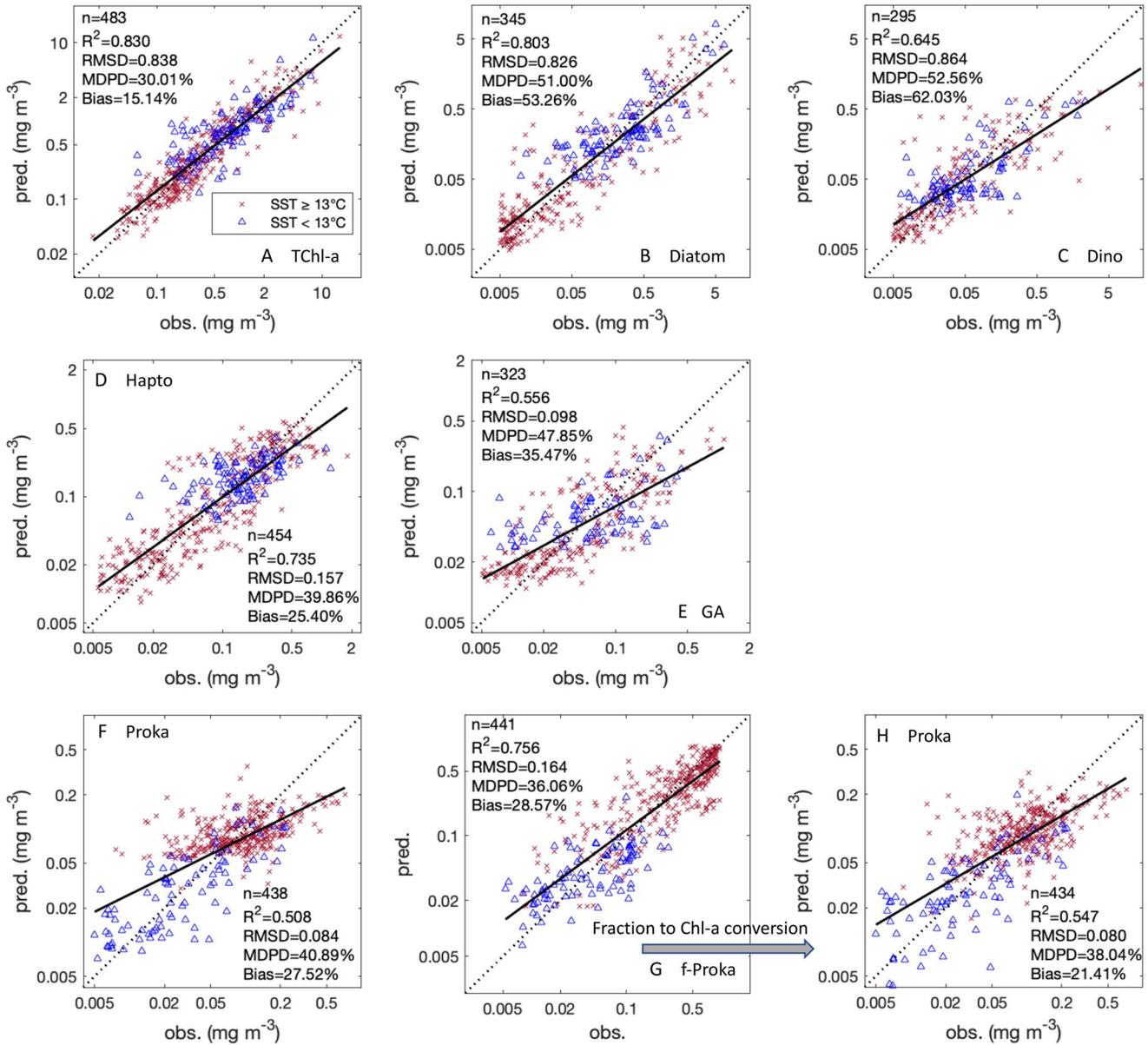
$$y(\mathbf{a}, \mathbf{u}(Rrs), SST) = \ln(C_p^{sat}) = a_o + a_1 u_1^{sat} + a_2 u_2^{sat} + \dots + a_n u_n^{sat} + a_{SST} SST \quad (5)$$

where  $\mathbf{a}$  represents all the model coefficients and  $\mathbf{u}$  represents all the EOF score vectors derived from R<sub>rs</sub> data. Each  $\mathbf{u}$  component can be expressed as:

$$u_i^{sat} = f(Rrs) \quad (6)$$



**Figure 5.** Regressions between observed (x-axis, obs.) and predicted (y-axis, pred.) PFT quantities using EOF-SST hybrid algorithm: (a) TChl-a, (b) diatoms, (c) dinoflagellates, (d) haptophytes, (e) green algae, (f) prokaryotes, (g) *f*-Prokaryotes, (h) fraction-derived prokaryote Chl-a, (i) *Prochlorococcus*, (j) *f*-*Prochlorococcus*, and (k) fraction-derived *Prochlorococcus* Chl-a. EOF, empirical orthogonal function; PFT, phytoplankton functional type; SST, sea surface temperature.



**Figure 6.** Combined regressions between observed (x-axis, obs.) and predicted (y-axis, pred.) PFT quantities from two sets of EOF-SST hybrid algorithms based on different SST ranges: (a) TChl-a; Chl-a of (b) diatoms, (c) dinoflagellates, (d) haptophytes, (e) green algae, and (f) prokaryotes; (g) *f*-Prokaryotes and (h) fraction-derived prokaryote Chl-a. The dotted black line shows the 1:1 line and the solid black line indicates the regression based on the whole data set. EOF, empirical orthogonal function; PFT, phytoplankton functional type; SST, sea surface temperature.

To estimate the final uncertainty of the retrieved PFTs,  $\sigma_y$ , we assume that the uncertainties due to  $\alpha$ ,  $u$ , and SST in Equation 5 are not correlated with each other. According to the Guide to Uncertainty in Measurement (JCGM, 2008), the combined uncertainty of different sources could be estimated based on the law of propagation of uncertainty. Using the partial differences, the uncertainty of the PFT retrievals is presented theoretically as:

$$\sigma_y = \sqrt{\sigma_{y(Rrs)}^2 + \sigma_{y(a)}^2 + \sigma_{y(SST)}^2} = \sqrt{\sum_{i=1}^N \left( \frac{\partial y}{\partial Rrs_i} \right)^2 \sigma_{Rrs_i}^2 + \sum_{i=0}^n \left( \frac{\partial y}{\partial a_i} \right)^2 \sigma_{a_i}^2 + \left( \frac{\partial y}{\partial SST} \right)^2 \sigma_{SST}^2}. \quad (7)$$

**Table 3**

*Absolute Uncertainties of  $R_{rs}$  ( $\sigma_{R_{rs}}$ ,  $Sr^{-1}$ ) for Different Sensors in the Merged Products Derived Based on Maritorena et al. (2010)*

| Wavebands (nm)            | 412     | 443     | 490     | 510     | 531     | 547     | 555     | 670      | 678      |
|---------------------------|---------|---------|---------|---------|---------|---------|---------|----------|----------|
| MODIS $\sigma_{R_{rs}}$   | 0.00071 | 0.00063 | 0.00049 | –       | 0.00024 | 0.00019 | –       | 0.000055 | 0.000030 |
| MERIS $\sigma_{R_{rs}}$   | 0.00066 | 0.00059 | 0.00047 | 0.00033 | –       | –       | 0.00023 | 0.00010  | 0.000098 |
| SeaWiFS $\sigma_{R_{rs}}$ | 0.00072 | 0.00064 | 0.00050 | 0.00036 | –       | –       | 0.00025 | 0.000075 | –        |
| RMS $\sigma_{R_{rs}}$     | 0.00070 | 0.00062 | 0.00049 | 0.00035 | 0.00024 | 0.00019 | 0.00024 | 0.000080 | 0.000072 |

*Note.* The root mean square (RMS) was taken as the uncertainty of the merged products.

Since the uncertainties propagated from errors of model parameters ( $\sigma_{y(a)}$ ) and SST ( $\sigma_{y(SST)}$ ) are both linear, they can be analytically derived and expressed together as:

$$\sigma_{y(a)}^2 + \sigma_{y(SST)}^2 = \sum_{i=1}^N (u_i^{sat})^2 \sigma_{a_i}^2 + SST^2 \sigma_{a_{SST}}^2 + (a_{sst})^2 \sigma_{SST}^2. \quad (8)$$

where  $\sigma_{SST} = 0.46$  °C (Worsfold et al. [2020] for the OSTIA SST product);  $\sigma_{a_i}$  and  $\sigma_{a_{SST}}$  were determined during the cross-validation procedure as described in Section 2.2.1. To further illustrate how  $\sigma_{a_i}$  and  $\sigma_{a_{SST}}$  were determined, Figure S4 shows the distributions of the coefficients derived from all 500 permutations in the cross-validation, using diatoms as an example. As the coefficient distributions followed the normal distribution, the uncertainty of each coefficient was determined by calculating the corresponding standard deviation. The uncertainties of the model coefficients for all other PFT quantities were also determined in the same manner.

Since the fraction-derived prokaryotes and *Prochlorococcus* Chl-a were calculated by multiplying their fractions and the TChl-a together, the corresponding uncertainties were determined by the uncertainties of TChl-a and that of *f*-Prokaryotes (*f-Prochlorococcus*). Using  $y_{proka\_conv}$ ,  $y_1$ , and  $y_2$  to denote the fraction-derived prokaryote Chl-a, TChl-a, and *f*-Prokaryotes, the uncertainty of the fraction-derived prokaryote Chl-a,  $\sigma_{y_{proka\_conv}}$ , can be determined by the uncertainty of TChl-a ( $\sigma_{y_1}$ ) and *f*-Prokaryotes ( $\sigma_{y_2}$ ) through the following equation:

$$\sigma_{y_{proka\_conv}} = \sqrt{\sigma_{y_1}^2 + \sigma_{y_2}^2 + 2\sigma_{y_1}\sigma_{y_2}r_{12}}, \quad (9)$$

where  $r_{12}$  is the correlation coefficient between  $\sigma_{y_1}$  and  $\sigma_{y_2}$  (as both depend on SST-related uncertainties). Equation 9 also applies for the uncertainty of the fraction-derived *Prochlorococcus* Chl-a.

Among all uncertainty components in Equation 7, the uncertainty propagated from errors in the satellite  $R_{rs}$  spectra,  $\sigma_{y(R_{rs})}$ , is the challenging part to quantify, as it is nonlinear and not as straightforward as the other two uncertainty sources, due to the EOF analysis performed with the spectra. We therefore used an MC simulation-based approach to estimate the  $\sigma_{y(R_{rs})}$  as detailed in the following section.

### 2.3.2. $R_{rs}$ Uncertainty Propagation

Based on the uncertainty of the water leaving radiance for SeaWiFS, MODIS-Aqua, and MERIS reported in Maritorena et al. (2010),  $R_{rs}$  absolute uncertainties for these sensors were derived and used in the GlobColour product. In our study, we took the root mean square (RMS) of the common bands from two or three sensors as the uncertainty of the merged products (Table 3). Using the matchup data for the merged products at nine bands, the following steps were carried out to fulfill the uncertainty propagation from  $R_{rs}$  to the PFTs.

**Table 4**

Statistics of Regression Models for TChl-a, Six PFT Chl-a, Fractions of Prokaryotes and Prochlorococcus and the Corresponding Fraction-Derived Chl-a Using SST and EOF Modes Based on the Nine-Band  $R_{rs}$  Matchups From Merged OC Products (Upper Panel)

|   | $N$ | MDPD (%)     | RMSD ( $\text{mg m}^{-3}$ ) | $R^2$       | MDPDcv(%)    | RMSDcv ( $\text{mg m}^{-3}$ ) | $R^2$ cv    |
|---|-----|--------------|-----------------------------|-------------|--------------|-------------------------------|-------------|
| EOF-SST hybrid model                    |     |              |                             |             |              |                               |             |
| TChl-a                                  | 483 | <b>31.55</b> | <b>1.08</b>                 | <b>0.82</b> | <b>31.51</b> | <b>1.14</b>                   | <b>0.81</b> |
| Diatoms                                 | 345 | <b>56.25</b> | <b>0.92</b>                 | <b>0.77</b> | <b>57.71</b> | <b>0.98</b>                   | <b>0.76</b> |
| Dinoflagellates                         | 295 | <b>54.00</b> | <b>0.89</b>                 | <b>0.62</b> | <b>54.68</b> | <b>0.71</b>                   | <b>0.60</b> |
| Haptophytes                             | 454 | <b>42.88</b> | <b>0.16</b>                 | <b>0.71</b> | <b>43.63</b> | <b>0.17</b>                   | <b>0.69</b> |
| Green algae                             | 323 | <b>51.57</b> | <b>0.10</b>                 | <b>0.53</b> | <b>52.96</b> | <b>0.11</b>                   | <b>0.51</b> |
| Prokaryotes                             | 438 | <b>43.79</b> | <b>0.09</b>                 | <b>0.42</b> | <b>45.44</b> | <b>0.09</b>                   | <b>0.38</b> |
| <i>f</i> -Prokaryotes                   | 441 | 47.35        | 0.21 <sup>a</sup>           | 0.70        | 47.72        | 0.21 <sup>a</sup>             | 0.68        |
| Fraction-derived prokaryotes            | 434 | <b>42.13</b> | <b>0.09</b>                 | <b>0.46</b> |              |                               |             |
| <i>Prochlorococcus</i>                  | 204 | 42.60        | <b>0.02</b>                 | <b>0.24</b> | 44.97        | <b>0.02</b>                   | 0.17        |
| <i>f-Prochlorococcus</i>                | 210 | 41.52        | 0.09 <sup>a</sup>           | 0.62        | 43.99        | 0.09 <sup>a</sup>             | 0.57        |
| Fraction-derived <i>Prochlorococcus</i> | 198 | 39.90        | <b>0.02</b>                 | <b>0.24</b> |              |                               |             |
| Original EOF-based algorithm            |     |              |                             |             |              |                               |             |
| TChl-a                                  | 394 | 37.41        | 1.24                        | 0.76        | 37.08        | 1.27                          | 0.75        |
| Diatoms                                 | 306 | 73.70        | 1.21                        | 0.65        | 74.74        | 1.29                          | 0.63        |
| Dinoflagellates                         | 272 | 55.32        | 0.93                        | 0.62        | 57.29        | 0.72                          | 0.59        |
| Haptophytes                             | 387 | 47.16        | 0.22                        | 0.64        | 48.62        | 0.24                          | 0.61        |
| Green algae                             | 262 | 55.81        | 0.11                        | 0.51        | 56.26        | 0.11                          | 0.48        |
| Prokaryotes                             | 367 | 53.70        | 0.13                        | 0.15        | 55.08        | 0.13                          | 0.11        |
| <i>Prochlorococcus</i>                  | 142 | 39.65        | 0.02                        | 0.24        | 42.68        | 0.02                          | 0.18        |

Note.  $N$  is the number of valid matchups for each parameter. Cross-validation was not applied for the fraction-derived prokaryotes and *Prochlorococcus* Chl-a because they are the results of the multiplication between their fractions and the TChl-a. As a comparison, the statistics of the previous EOF-based algorithm (without SST) by Xi et al. (2020) for the TChl-a and six PFT Chl-a are also presented (lower panel). Bold marks the improved (or same) statistics.

Abbreviations: EOF, empirical orthogonal function; MDPD, median percent difference; OC, ocean color; PFT, phytoplankton functional type; RMSD, root-mean-square difference; SST, sea surface temperature.

<sup>a</sup>The unit ( $\text{mg m}^{-3}$ ) of RMSD does not apply to *f*-Prokaryotes and *f-Prochlorococcus* which have a range of 0–1.

1. The 483  $R_{rs}$  matchup spectra were randomly divided equally into two data sets—50% as the training data set (in a total of 242), and the other 50% as the testing data set (in a total of 241). The corresponding matchups of in situ PFT and retrieved PFT data were also divided accordingly (Table 2).
2. For the training data sets, we performed 10,000 MC simulations to randomly introduce for each band the  $R_{rs}$  uncertainty ( $\sigma_{Rrs}$ ) to each spectrum in the training data set (in a total of 2,420,000 simulated spectra).
3. The MC simulated  $R_{rs}$  spectra were applied to the EOF-SST hybrid algorithm to estimate the PFTs with  $R_{rs}$  uncertainty taken into account. For each sample, 10,000 estimates of the PFT were generated from the 10,000 MC simulated  $R_{rs}$ , so that the uncertainty (standard deviation,  $\sigma_{MC-PFT}$ ) of the PFTs were determined based on these 10,000 estimates for each sample.
4. When the  $\sigma_{MC-PFT}$  for all samples in the training data set were determined through Step 3, a look-up table (LUT) was built for each PFT by fitting  $\sigma_{MC-PFT}$  as a function of the retrieved PFT.
5. The LUT for each PFT was applied to the testing data set for the uncertainty validation, and also to the satellite PFT products to derive per-pixel uncertainty of the satellite PFT due to  $R_{rs}$ ,  $\sigma_{y(Rrs)}$ , which was combined with uncertainties from the other sources via Equation 7 to derive the final uncertainty of PFT satellite retrievals.

### 2.3.3. Assessment of the Per-Pixel PFT Uncertainty

With the steps in Section 2.3.2, the uncertainty propagated from the  $R_{rs}$  to the satellite retrieved PFTs (the term  $\sigma_{y(Rrs)}^2$  in Equation 7) was determined by applying the LUT to each pixel of the satellite-derived PFT products. Together with the other two terms,  $\sigma_{y(a)}^2$  and  $\sigma_{y(SST)}^2$  which were calculated analytically through Equation 8, the combined PFT uncertainty  $\sigma_y$  of each pixel from different sources was ultimately obtained.

## 3. Results and Discussion

### 3.1. EOF-SST Hybrid Algorithms for PFT Retrievals

#### 3.1.1. EOF-SST Hybrid Algorithm Based on the Whole Matchup Data Set

Before setting up the EOF-SST algorithm, we first updated the original SST-independent EOF-based algorithm of Xi et al. (2020) using the larger input data set used in this study. The performance of the updated algorithm showed nearly identical performance as compared to the original one presented in Xi et al. (2020) (details not shown), suggesting that the original algorithm can be hardly improved by purely enlarging the training data set. We then applied the EOF-SST hybrid algorithm developed in the current study to derive the TChl-a, PFT Chl-a, and the fractions of two PFTs based on the regression models built using the EOF scores derived from the nine-band  $R_{rs}$  data, SST, and the in situ PFT data. As presented in Table 4 and Figures 5a–5f and 5h, compared to the original algorithm in Xi et al. (2020), the EOF-SST hybrid algorithm shows significant improvements for all predicted quantities except for *Prochlorococcus* where weak performance still remains. For TChl-a and Chl-a of diatoms, haptophytes, dinoflagellates, and green algae,  $R^2$  increased from a range of 0.51–0.76 to a range of 0.53–0.82 and  $R^2_{cv}$  increased from a range of 0.48–0.75 to a range of 0.51–0.81. MDPD reduced from its range in Xi et al. (2020) of 37%–74% to a range of 32%–56%, while MDPD<sub>cv</sub> reduced from 37%–75% to a range of 32%–58%. RMSD and RMSD<sub>cv</sub> values are also significantly lower in the EOF-SST hybrid algorithm compared to the previous results.

To further improve the prediction of prokaryote and *Prochlorococcus* Chl-a, the hybrid algorithm was also trained to retrieve the fractions of prokaryotes and *Prochlorococcus* to TChl-a. This was motivated because prokaryotes dominate in low TChl-a midlatitude to low-latitude waters, so generally their Chl-a is low. By using their fraction instead of Chl-a, a better spread of the data is achieved that enhances the signal to be retrieved which is beneficial for application in abundance-based PSC retrievals (e.g., Brewin et al., 2010). As expected, the prediction models for the two fractions performed well with  $R^2 > 0.62$  and MDPD within 47% (Table 4). Though the overall performance of the fraction retrieval for the two PFTs had been improved, the regression between the predicted and observed *f-Prochlorococcus* (Figure 5j) show that higher discrepancies exist in low fraction values, indicating that it is still difficult to accurately derive *Prochlorococcus* Chl-a from the fraction when *f-Prochlorococcus* < 0.2. Using predicted TChl-a, the fractions were further converted to Chl-a for the two PFTs. Table 4 shows that the fraction-derived prokaryote Chl-a retrieval displays improved performance compared to the directly retrieved prokaryotes ( $R^2$  of 0.46 vs. 0.42, MDPD of 42% vs. 44%, and with equivalent RMSD of 0.08 mg m<sup>-3</sup>), but is downgraded compared to the *f-Prokaryotes* (Figure 5g). The fraction to Chl-a conversion scheme shows little improvement in predicting *Prochlorococcus* Chl-a (Table 4 and Figure 5k vs. Figure 5i). Though *f-Prochlorococcus* is overall better predicted compared to the direct retrieval of *Prochlorococcus* Chl-a, by using the conversion,  $R^2$ , and RMSD of the fraction-derived *Prochlorococcus* Chl-a estimation deteriorate due to the low signal-to-noise ratio in the retrieved TChl-a and *f-Prochlorococcus*. The MDPD and bias are however slightly improved. Weak prediction performance of the fraction-derived *Prochlorococcus* Chl-a reveals that it is still challenging to enhance their retrieval accuracy to the same level as other PFTs due to the low concentrations and small variability (Xi et al., 2020).

#### 3.1.2. SST-Separated Hybrid Algorithms for Different SST Regimes

As described in Section 2.2.2, SST-separated hybrid algorithms were developed to retrieve the PFT quantities respectively for the two temperature regimes. Table S2 summarizes the coefficients fitted in the stepwise regression models based on the whole data set, data set with SST ≥ 13 °C, and data set with SST < 13 °C, respectively. EOF modes chosen for different PFTs vary with different data sets, such that SST as an additional regression term may not always been used in the final prediction models due to different levels of

**Table 5**  
Combined Statistics of the Regression Models From the SST-Separated Hybrid Algorithms for Matchup Data With SST  $\geq 13$  °C and With SST  $< 13$  °C

|                              | <i>N</i> | $R^2$       | RMSD (mg m <sup>-3</sup> ) | MDPD (%)     |
|------------------------------|----------|-------------|----------------------------|--------------|
| TChl-a                       | 483      | <b>0.83</b> | <b>0.84</b>                | <b>30.01</b> |
| Diatoms                      | 345      | <b>0.82</b> | <b>0.83</b>                | <b>51.00</b> |
| Dinoflagellates              | 295      | <b>0.65</b> | <b>0.86</b>                | <b>52.56</b> |
| Haptophytes                  | 454      | <b>0.74</b> | <b>0.16</b>                | <b>39.86</b> |
| Green algae                  | 323      | <b>0.56</b> | <b>0.10</b>                | <b>47.85</b> |
| Prokaryotes                  | 438      | <b>0.51</b> | <b>0.08</b>                | <b>40.89</b> |
| <i>f</i> -Prokaryotes        | 441      | <b>0.76</b> | <b>0.16<sup>a</sup></b>    | <b>36.06</b> |
| Fraction-derived prokaryotes | 434      | <b>0.55</b> | <b>0.08</b>                | <b>38.04</b> |

Note. Improved (or same) parameters are marked as bold, by comparing to those from the hybrid algorithm without separating SST (Table 3).

Abbreviations: MPDP, median percent difference; RMSD, root-mean-square difference; SST, sea surface temperature.

<sup>a</sup>Note that the unit (mg m<sup>-3</sup>) of RMSD does not apply to *f*-Prokaryotes.

correlation with PFTs in different SST regimes. The SST term had been identified as insignificant within the stepwise minimization method routine performed in the model regression procedure if the contribution of SST to the PFT estimation was neglectable. Moreover, the weighting coefficient fitted on the SST term changed when different data sets were used. For instance,  $a_{SST}$  fitted in the prediction models for prokaryote Chl-a using the whole data set was 0.088, and increased to 0.117 for the data set with SST  $< 13$  °C, but was not used in the prediction model for the data set with SST  $\geq 13$  °C which is also consistent with the SST-PFT relationship (Figure 3). The 10-point running mean trend showed that SST had a distinct positive correlation with the prokaryote Chl-a but the correlation turned insignificant when SST was higher than 13 °C.

Table 5 shows the combined statistics of the regression models from SST-separated hybrid algorithms for the two temperature regimes, while Table S3 displays separately the statistics for the two regimes. Generally, Table S3 shows that TChl-a and the eukaryotic PFT Chl-a are better retrieved for the data set with SST  $\geq 13$  °C though RMSD is higher simply because their Chl-a spanned a wider range than in cold waters (Figure 6). In contrast, both direct and fraction-derived prokaryote Chl-a retrievals show significantly improved performance for the data set with SST  $< 13$  °C, suggesting that better estimations for prokaryotes can be achieved by considering the strong correlation between prokaryote Chl-a and SST in cold waters. For a global PFT retrieval scheme in this study, we focused

more on the combined statistics and the overall performance of the SST-separated hybrid algorithms. As shown in Table 5 and Figure 6, the improvement for TChl-a estimated from the SST-separated algorithms is rather small, indicating that responses in the TChl-a concentration to different SST regimes are relatively stable. For predictions of Chl-a of all PFTs, except *Prochlorococcus* for which the separation of SST does not apply, the SST-separated algorithms perform considerably better, indicated by improved statistics in terms of  $R^2$ , RMSD, and MDPD (Table 5). Prokaryotes, both in terms of Chl-a and fraction, show the most promising improvement compared to that from the non-separated EOF-SST hybrid algorithm. With good performances in retrieving the *f*-prokaryotes and TChl-a, the prokaryote Chl-a is more accurately derived by converting the fraction to concentrations (Figure 6h) compared to the directly retrieved prokaryote Chl-a (Figure 6f), with an increase of  $R^2$  from 0.51 to 0.55 and reduced MDPD from 40.89% to 38.04%. Though prokaryotes retrieval is still not as good as the other PFTs such as diatoms and haptophytes, it is equivalent to the green algae retrieval with even lower MDPD. Compared to the prokaryote retrieval in Xi et al. (2020), this result is a significant improvement after a series of experiments by including SST in the retrieval model, establishing separated models based on SST regimes, and retrieving first the fraction and performing the conversion.

Our global in situ PFT data with SST around the threshold 13 °C located roughly between 35 and 45°N or °S, correspond well to the findings of Ward (2015): Their SST-dependent PSC models show a general decline in picophytoplankton biomass polewards of 40°N or °S, contrary to the overall trends for TChl-a, microphytoplankton, and nanophytoplankton biomass. Our Figure 3 shows the same declining trend for prokaryote (picophytoplankton) Chl-a in high latitude waters with SST  $< 13$  °C, while Chl-a of diatoms, dinoflagellates (corresponding to microphytoplankton), and haptophytes (corresponding to nanophytoplankton) in cold waters are generally higher than in waters of SST  $\geq 13$  °C. Similarly, Brewin, Ciavatta, et al. (2017) chose a relatively higher SST value (15 °C) to partition their data set into two subsets with roughly equal numbers in the North Atlantic Ocean, and also found significantly different model parameters for high and low temperatures. Coincidentally, Moore & Brown (2020) analyzed the trends in parameters of PSC fraction models for the whole Atlantic Ocean, and found that the different model parameter changes occurred in the SST ranges of 5 °C–15 °C, 15 °C–25 °C, and  $> 25$  °C, respectively. All these similar behaviors in models incorporating SST for PFT or PSC estimations suggest ecological connections. Mouw et al. (2019) have investigated the ecological drivers for phytoplankton size distribution. Among the nine considered environmental parameters (e.g., eutrophic depth, TChl-a, SST, mixed layer depth, etc.), they found that the euphotic depth

is the most important parameter driving the variability of percent microphytoplankton ( $S_{\text{fm}}$ ) on a global scale, covering the central gyres and the majority of the high-latitude ocean. However, in other regions,  $S_{\text{fm}}$  distribution is driven by a balance of light and mode of nutrient delivery. In summary, SST and other environmental parameters always covary with each other and have compound effect on PFT abundance and variability, and different PFTs behave differently both in time and space. Though it is still challenging to clearly explain the ecological connections behind these phenomena, a number of indications provided by the studies above lead to some potential directions to be investigated in future.

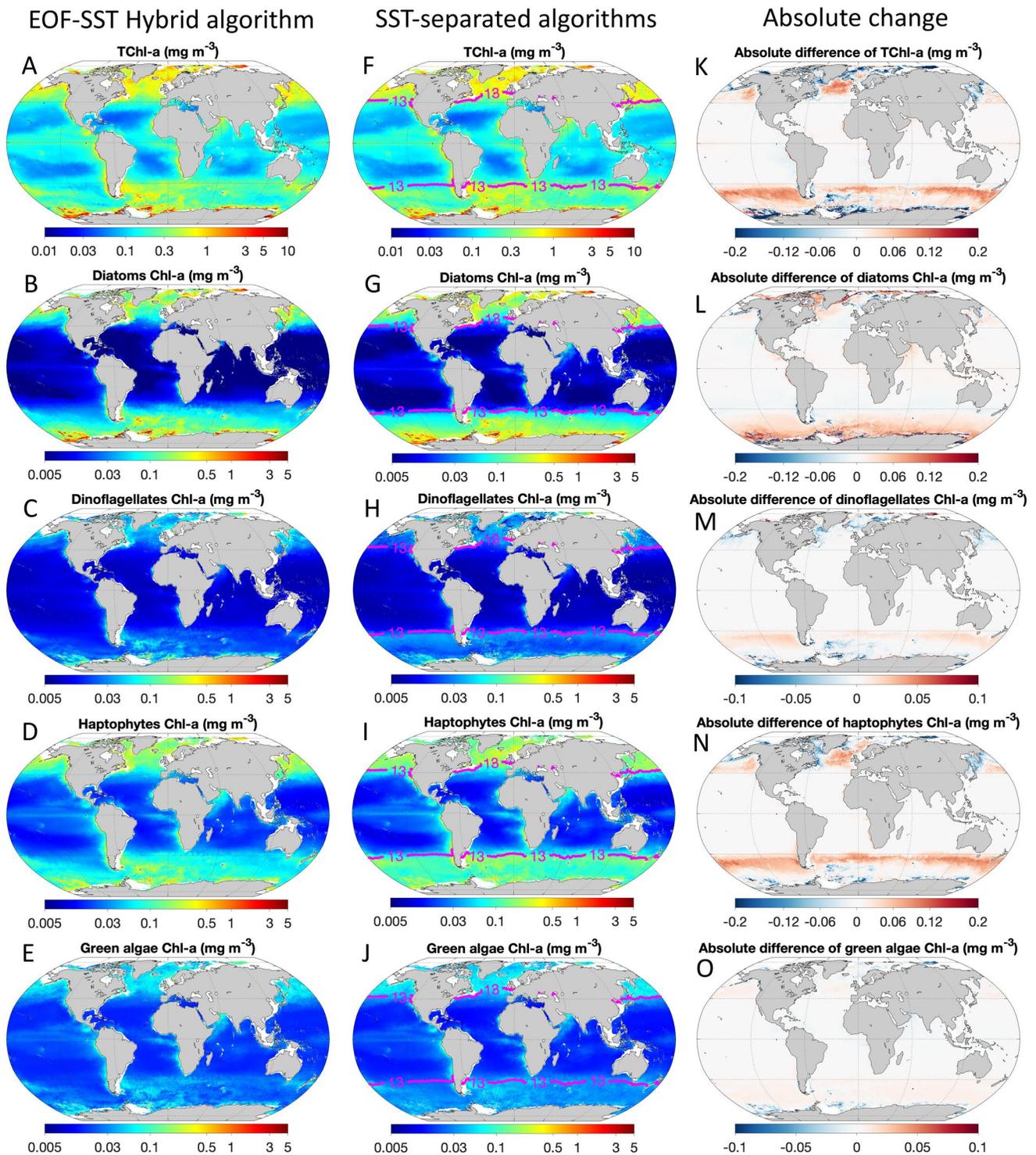
### 3.2. Global Maps of PFT Quantities From Merged $R_{\text{rs}}$ Products

The improved EOF-SST hybrid algorithms from both Sections 3.1.1 and 3.1.2 were applied to the merged  $R_{\text{rs}}$  products and SST data to derive the global TChl-a, PFT Chl-a, and the fractions (see Figure 4, part model application). To illustrate the global distribution of the PFTs, Figures 7 and 8 show the annual mean generated from the derived monthly PFT quantities (except *Prochlorococcus*) for the year 2011 using the EOF-SST hybrid algorithm (established in Section 2.2.1) and the SST-separated algorithms (established in Section 2.2.2), respectively, with the absolute difference between the two products. Since *Prochlorococcus* barely exist at cold temperatures (see Section 2.2), global maps of *Prochlorococcus* Chl-a and *f-Prochlorococcus* are generated only for regions with  $SST \geq 10^\circ\text{C}$  (Figure 9).

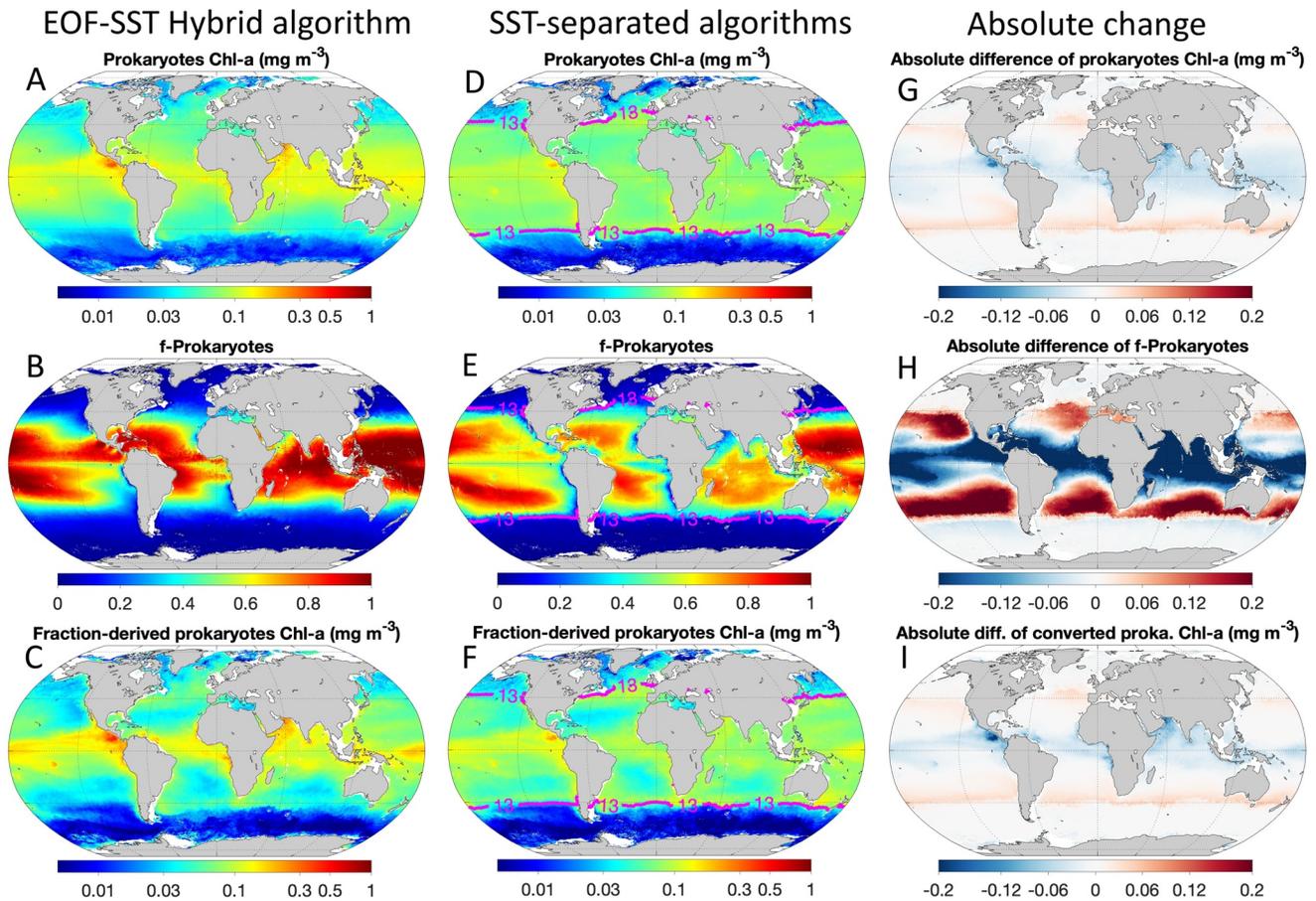
In general, distribution patterns of the retrieved TChl-a and the four eukaryotic PFTs from the EOF-SST hybrid algorithm are consistent with those from the combination of the SST-separated algorithms (Figure 7). However, distinct differences between the retrievals from the two approaches are found in latitudes higher than  $40^\circ\text{N/S}$  (Figures 7k–7o). Compared to TChl-a derived from the EOF-SST hybrid algorithm (Figure 7a), TChl-a derived from SST-separated algorithms (Figure 7f) are elevated in the North Pacific and Atlantic Oceans, as well as in most regions between  $40^\circ\text{S}$  and  $60^\circ\text{S}$  of the Southern Ocean, while decreased in most of the high latitude marginal seas and in the Arctic except for the Barents Sea (Figure 7k). Diatom Chl-a from the SST-separated approach is enhanced in high latitudes and coastal waters (Figure 7g vs. Figure 7b). Haptophyte Chl-a is enhanced mostly in the moderately high latitudes (e.g.,  $45^\circ\text{S}$ – $60^\circ\text{S}$  near-polar fronts) and decreased in very high latitudes and marginal seas using the SST-separated algorithms (Figure 7n). Changes in Chl-a for dinoflagellates and green algae generally follow the changes in haptophyte Chl-a but are much milder (Figures 7m and 7o). Compared to the retrievals of Xi et al. (2020), satellite-derived PFTs from the SST-separated algorithms are in better agreement with the equivalent products from other global studies (e.g., Brewin et al., 2015; Hirata et al., 2011; Losa et al., 2017).

The retrievals of prokaryote Chl-a, *f*-prokaryotes and the fraction-derived prokaryote Chl-a using SST-separated algorithms present generally lower values globally except in the regions around  $20^\circ\text{S}$ – $40^\circ\text{S}$  in the southern hemisphere,  $20^\circ\text{N}$ – $40^\circ\text{N}$  in the north Pacific Ocean and between  $30^\circ\text{N}$  and  $45^\circ\text{N}$  in the north Atlantic Ocean (Figures 8g–8i). The fraction-derived Chl-a shows more reasonable global distribution (Figures 8c and 8f) compared to the direct retrievals (Figures 8a and 8d), given the improved performance via fraction conversion for prokaryotes in Section 3.1. The EOF-SST hybrid algorithm derived *f*-prokaryotes is saturated (up to 1) in the gyres or their adjacent regions (Figure 8b). This saturation is remarkably reduced by the SST-separated algorithms (Figure 8e) thanks to the better description of the prokaryotes' dependency on the temperature in the algorithm, resulting in that fraction-derived prokaryote Chl-a from SST-separated algorithms (Figure 8f) has the most reliable retrieval quality among all prokaryote Chl-a retrievals shown in Figure 8. Compared to the prokaryote retrieval in Xi et al. (2020), where the original EOF-based algorithm overestimates the prokaryote Chl-a dramatically, the fraction-derived prokaryote Chl-a in this study shows better agreements with previous studies (Hirata et al., 2011; Losa et al., 2017) but is relatively lower in the polar regions.

Regarding *Prochlorococcus*, although the fraction-derived Chl-a showed no distinct improvement in model performance compared to the direct Chl-a retrieval (Section 3.1.1), the global retrieval depicts an overall decrease in the fraction-derived Chl-a (Figure 9c) compared to the direct retrieval (Figure 9a), and is in closer agreement with in situ observations and other satellite retrievals (Alvain et al., 2008; Hirata et al., 2011). The conversion to Chl-a is restricted by the high uncertainty at low TChl-a and also the low variability of *Prochlorococcus* Chl-a (Xi et al., 2020).

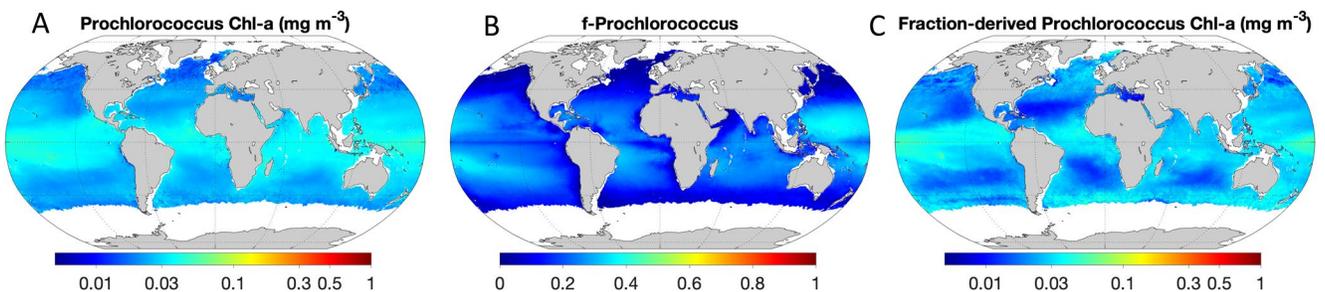


**Figure 7.** Satellite-derived estimates of annual (2011) mean surface TChl-a, Chl-a of diatoms, dinoflagellates, haptophytes, and green algae. Panels (a–e): EOF-SST hybrid algorithm with non-separated SST. Panels (f–j): Combined estimates from SST-separated hybrid algorithms for SST  $\geq 13$  °C and SST  $< 13$  °C, respectively. The magenta curve indicates the isotherm of 13 °C. Panels (k–o): Absolute difference between the combined estimates from SST-separated algorithms and that from EOF-SST hybrid algorithm, that is, panels (f–j) minus panels (a–e). EOF, empirical orthogonal function; PFT, phytoplankton functional type; SST, sea surface temperature.



**Figure 8.** Same as in Figure 7 but for annual (2011) mean surface Chl-a of prokaryotes, *f*-Prokaryotes, and the fraction-derived prokaryote Chl-a.

The general geographic patterns of Chl-a for the six PFTs, using the algorithms proposed in this study, are consistent with current knowledge (e.g., Bracher et al., 2020; Brewin et al., 2015; Hirata et al., 2011; Losa et al., 2017). It also indicates that spectral-based approaches have the capability to derive PFT information equivalently to or better than abundance-based approaches. Though TChl-a has been considered as an effective first-order predictor of absorption spectra (Cael et al., 2020), PFTs that have different assemblages of accessory pigments, in addition to chlorophyll *a*, are extractable from spectral measurements with the aid of environmental parameters. The improvements obtained in our study also agree well with previous studies that incorporated SST to their PSC models. Ward (2015) reported increases of microphytoplankton and nanophytoplankton Chl-a and decreases of picophytoplankton Chl-a in cold polar regions using the temperature-dependent functions compared to the temperature-independent functions. Moore & Brown (2020),



**Figure 9.** Satellite-derived estimates of annual (2011) mean surface (a) Chl-a of *Prochlorococcus*, (b) *f*-*Prochlorococcus*, and (c) fraction-derived *Prochlorococcus* Chl-a using the EOF-SST hybrid algorithm.

though only focused on the temperate to tropical zones of the Atlantic Ocean, noted that picophytoplankton abundance estimated by the SST-included models was lower in the gyres and temperate zones and higher in equatorial region than that from the baseline model without SST incorporation. These changes are also observed in the PFTs retrieved by our EOF-SST hybrid algorithms. For instance, compared to our original algorithm in Xi et al. (2020), the EOF-SST hybrid algorithms show much more realistic distributions: they generate higher Chl-a in polar waters and lower Chl-a in midlatitude to low-latitude for diatoms and haptophytes; much lower prokaryote Chl-a in polar waters and gyres, and higher prokaryote Chl-a in equatorial region. The prokaryotic phytoplankton distributions particularly show more contrasting patterns between gyre and non-gyre regions. Our improved retrievals show higher TChl-a in the Southern Ocean where satellite derivations from conventional algorithms are often underestimated (e.g., Johnson et al., 2013; Soppa et al., 2014). PFT quantities are better retrieved by the SST-separated algorithms which account for the effect of SST more accurately in the statistical models by separating cold and warm waters than the hybrid algorithm, suggesting that subtle model configurations of such environmental parameters help to describe phytoplankton community variability more precisely. Nevertheless, validation with in situ pigment and/or PFT data is always favorable when more measurements are available.

### 3.3. PFT Uncertainty

#### 3.3.1. Look-up Table for Uncertainty due to $R_{rs}$

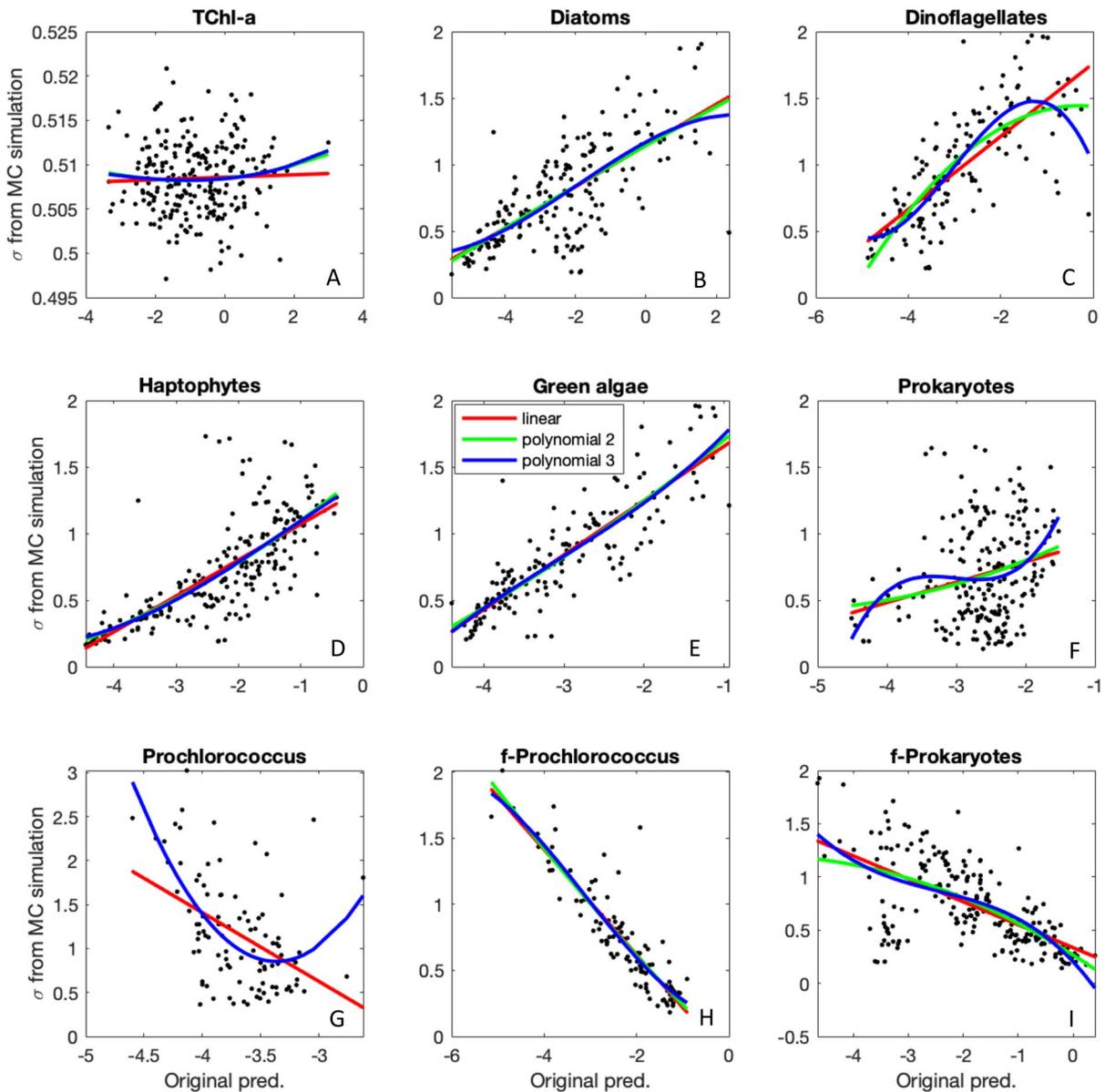
Following the steps listed in Section 2.3.2 to build the LUT for quantifying PFT uncertainty propagated from  $R_{rs}$  uncertainty, Figure 10 and Table 6 show the regressions and the statistical results of  $\sigma_{MC-PFT}$  against the originally predicted PFT quantities. Higher  $R^2$  and lower RMSD are achieved in general when higher degree of polynomial is used (Table 6). However, the difference between different regressions is rather small except for dinoflagellates, which is caused by their few data points at higher concentrations with corresponding lower MC-derived uncertainties (Figure 10c). Distinct differences between polynomial regressions also exist in those PFTs whose correlations to  $\sigma_{MC-PFT}$  are weak (e.g., Figures 10a, 10f and 10g). To be conservative, linear regressions were taken as the final LUT functions to determine  $\sigma_{y(Rrs)}$ .

It is noted that  $\sigma_{y(Rrs)}$  can not always be well defined by fitting a linear function. For example, the uncertainty from MC simulation for TChl-a shows little variation (0.495–0.525) (Figure 10a), indicating that the uncertainty is not dependent on TChl-a and does not change much with TChl-a ( $R^2 = 0$ ). Uncertainties associated with prokaryotic phytoplankton (prokaryotes and *Prochlorococcus*) Chl-a are not well correlated to their retrievals either ( $R^2 < 0.21$ ). The uncertainties of their fractions are however highly dependent on the fraction retrievals with an inverse correlation ( $R^2 > 0.40$ ); the retrieval of higher fractions thus bears lower uncertainty. The derived regressions imply that the LUTs quantify well the PFT uncertainty propagated from  $R_{rs}$  uncertainty for the non-prokaryotic PFT Chl-a and the fractions of the prokaryotic phytoplankton but not for the prokaryotic Chl-a. TChl-a uncertainty is relatively stable and not related to the retrieved TChl-a.

#### 3.3.2. Validation of PFT Uncertainty

The linear LUTs were applied to the retrieved PFT quantities in the testing data set to determine the corresponding  $\sigma_{y(Rrs)}$ . The final consolidated uncertainties of the retrieved PFTs from the testing data set,  $\sigma_y$ , were then estimated using Equation 5. With matchup data from the testing data set (Step 1 of Section 2.3.2), it is possible to assess whether or not the estimated uncertainties for the PFT products are accurate by comparing them to the actual error,  $\delta_y$ , defined as  $\delta_y = \ln(c_p) - \ln(c_o)$ . If the uncertainty  $\sigma_y$  is truly representative of its standard deviation and thus is reliable, the distribution of the actual errors normalized by the estimated errors  $\delta_y/\sigma_y$  should, to some extent, follow a standard centered normal distribution (Maritorena et al., 2010). Therefore, to validate the estimated PFT uncertainty  $\sigma_y$ , the testing data set was compared to the corresponding  $\delta_y$ . Note that  $\sigma_y$  and  $\delta_y$  are both natural-logarithmic based.

Figure 11 shows the histograms of  $\delta_y/\sigma_y$  distribution derived from the testing data set for all PFT quantities. The corresponding normal distributions determined by the mean and SD are also displayed in comparison with the centered standard normal distribution. For the majority of PFT quantities, the  $\delta_y/\sigma_y$  distribution coincides well with the standard normal distribution, with mean values close to zero and the SD vary-



**Figure 10.** Scatterplots of  $\sigma_{\text{PFT}}$  based on the MC simulations versus originally retrieved (natural-logarithmic based) PFTs. Regression lines of linear (red), polynomial fittings with degrees of 2 (green) and 3 (blue) are also shown. MC, Monte Carlo; PFT, phytoplankton functional type.

ing from 0.84 to 1.22. Diatoms, haptophytes, and green algae present slightly lower modeled uncertainty compared to the actual errors ( $SD > 1.09$ ). Relatively lower  $SD (< 0.8)$  are found for *Prochlorococcus* Chl-a and fraction, and the fraction-derived Chl-a both for prokaryotes and *Prochlorococcus*. The fraction-derived prokaryote Chl-a presents higher modeled uncertainty (i.e., lower  $SD$  of  $\delta_{\text{PFT}}/\sigma_{\text{PFT}}$  in Figure 11h) compared to the direct retrieval of prokaryote Chl-a (Figure 11f), even though the former shows better prediction performance (Table 3). This result suggests the possible underestimation of the actual errors in the direct retrieval for prokaryotes. *Prochlorococcus* Chl-a and fraction show overall higher modeled uncertainties compared to the actual error ( $SD \leq 0.73$ , Figures 11i–11k). However, a skewed distribution of  $\delta_y/\sigma_y$  is found for the direct retrieval of *Prochlorococcus* Chl-a and *f-Prochlorococcus* (Figures 11i and 11j) whereas the fraction-derived Chl-a shows milder skewness in their  $\delta_y/\sigma_y$  (Figure 11k). This result suggests the modeled uncertainty is better described for the fraction-derived *Prochlorococcus* Chl-a than the direct Chl-a retrievals. The validation of the PFT uncertainty indicates that our modeled uncertainty is in general close to the

**Table 6**  
*R<sup>2</sup> and RMSD of the Regression Functions Fitting the Relationship Between  $\sigma_{PFT}$  and Retrieved PFT Quantities (Using EOF-SST Hybrid Algorithm)*

|                              | Linear                |       | Polynomial 2          |       | Polynomial 3          |       |
|------------------------------|-----------------------|-------|-----------------------|-------|-----------------------|-------|
|                              | <i>R</i> <sup>2</sup> | RMSD  | <i>R</i> <sup>2</sup> | RMSD  | <i>R</i> <sup>2</sup> | RMSD  |
| Diatom Chl-a                 | 0.51                  | 0.262 | 0.51                  | 0.262 | 0.52                  | 0.262 |
| Dinoflagellate Chl-a         | 0.58                  | 0.277 | 0.62                  | 0.264 | 0.65                  | 0.252 |
| Green algae Chl-a            | 0.74                  | 0.208 | 0.74                  | 0.207 | 0.74                  | 0.207 |
| Haptophyte Chl-a             | 0.57                  | 0.244 | 0.57                  | 0.243 | 0.57                  | 0.243 |
| <i>Prochlorococcus</i> Chl-a | 0.21                  | 0.549 | 0.36                  | 0.494 | 0.36                  | 0.494 |
| Prokaryote Chl-a             | 0.07                  | 0.354 | 0.07                  | 0.354 | 0.10                  | 0.347 |
| TChl-a                       | 0.00                  | 0.004 | 0.01                  | 0.004 | 0.01                  | 0.004 |
| <i>f-Prochlorococcus</i>     | 0.81                  | 0.176 | 0.82                  | 0.175 | 0.82                  | 0.175 |
| <i>f-Prokaryotes</i>         | 0.40                  | 0.302 | 0.41                  | 0.30  | 0.42                  | 0.296 |

*Note.* Linear regressions were chosen for uncertainty quantification.

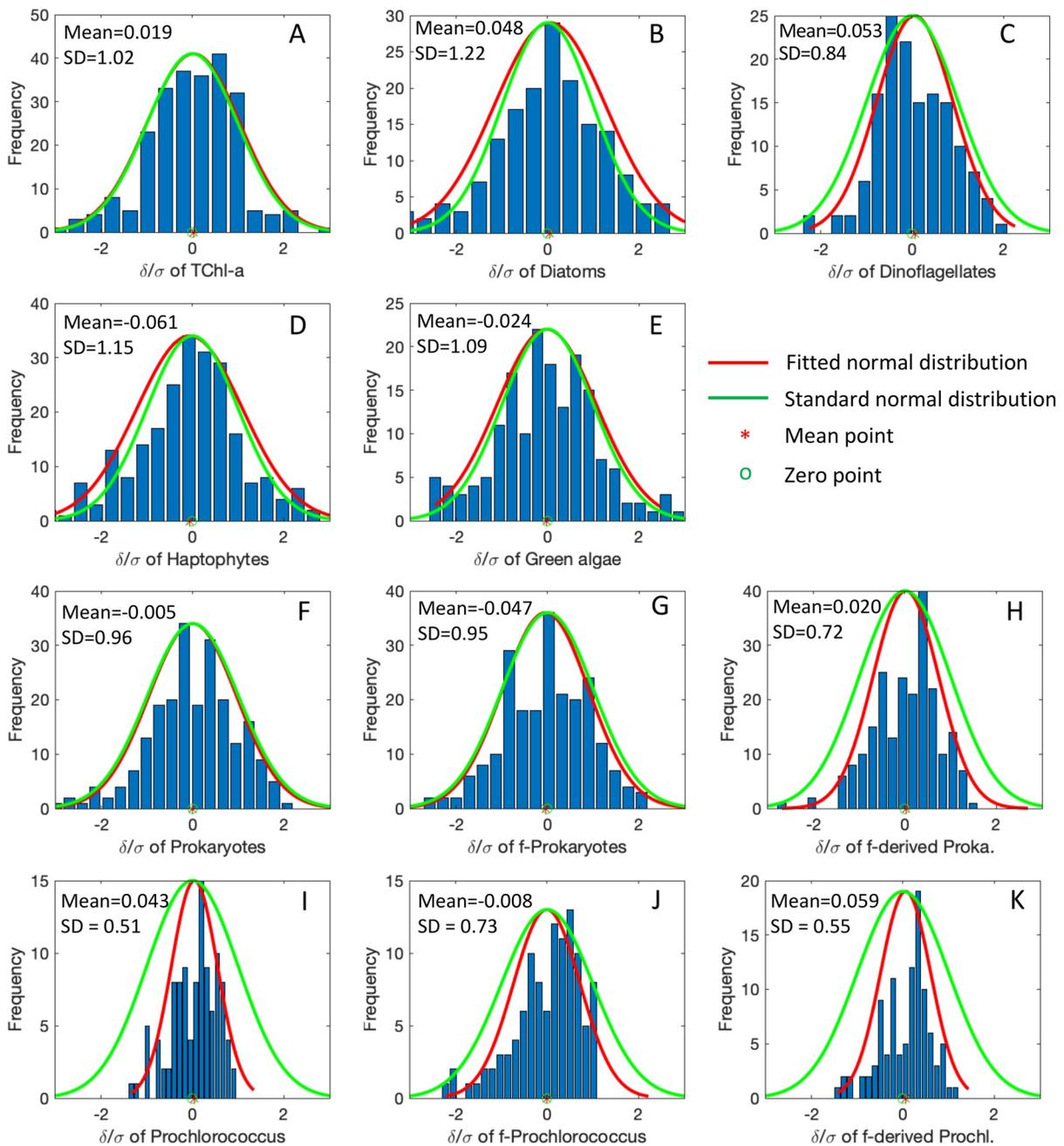
Abbreviations: EOF, empirical orthogonal function; PFT, phytoplankton functional type; RMSD, root-mean-square difference; SST, sea surface temperature.

actual error (*SD* of  $\delta_y/\sigma_y$  is close to 1), implying that the uncertainty assessment we performed in this study is reliable and conservative.

### 3.3.3. Per-Pixel PFT Uncertainty

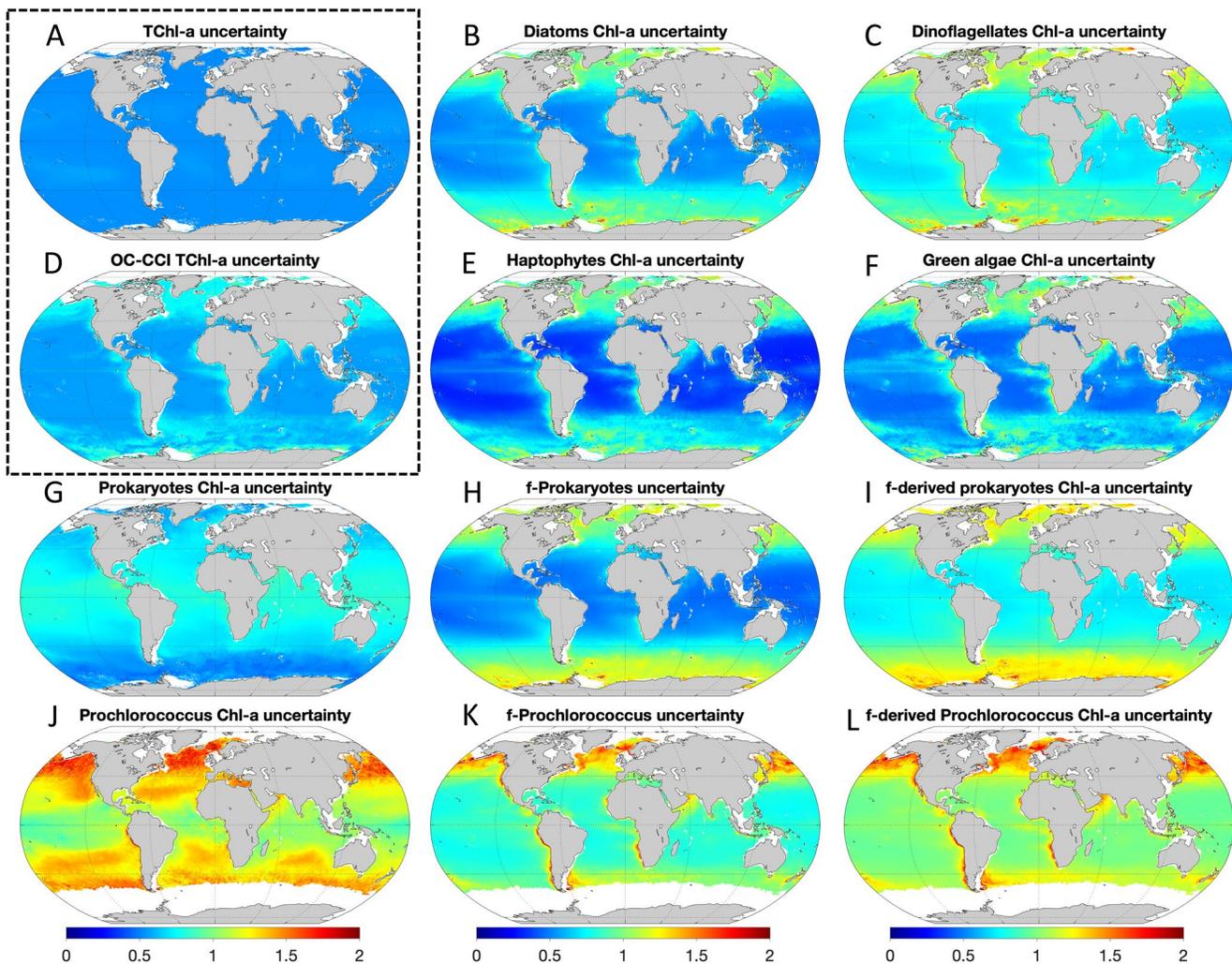
Satellite PFT uncertainties were generated for each pixel by applying all uncertainty terms in Equation 7 to the monthly retrieved PFT products. The annual mean uncertainty for the year 2011 was determined to be consistent with the global PFT maps in Figures 7–9. As a composite product, the annual mean uncertainty was obtained by computing the root mean square of the monthly uncertainty, the same as for the composite uncertainty of other OC products, for example, the ESA Ocean-Color Climate Change Initiative (OC-CCI) TChl-a product (Sathyendranath et al., 2019). It should be noted that in the ideal case, we should apply the uncertainty quantification scheme first to the PFT daily products, and then compute the monthly or yearly composite based on the multi-day uncertainties. However, detailed computations have shown that the monthly composites from the daily products have poorer spatial coverage compared to the directly derived monthly products, mainly due to fewer valid pixels for the nine-band  $R_{rs}$  spectra in the daily to monthly composites than from the direct monthly products. In addition, the uncertainty derived directly from the monthly products is very comparable to the monthly composite generated based on daily uncertainty (e.g., for TChl-a uncertainty the  $R^2 = 0.98$  and slope = 1.00 between the two derivations, details not shown). Therefore, in the present study, we applied the uncertainty quantification scheme directly to the monthly PFT products to save computing time and achieve better spatial coverage for the derived per-pixel uncertainties.

Figure 12 shows the annual composite of uncertainties for all the PFT quantities, including also the uncertainty for TChl-a derived from the EOF-SST hybrid algorithm in comparison to the uncertainty of the OC-CCI v4.2 TChl-a product (Figure 12a vs. Figure 12d). Uncertainties for diatoms (Figure 12b), dinoflagellates (Figure 12c), haptophytes (Figure 12e), and green algae (Figure 12f) display similar distribution patterns with their Chl-a retrievals, with low uncertainties in the gyres, and higher uncertainties in high latitude regions and marginal seas. The overall lowest uncertainty is obtained for haptophyte Chl-a (the natural logarithmic uncertainty varied from 0.28 to 1.35). The uncertainty for green algae Chl-a is also low (0.42–1.45), whereas diatom Chl-a shows higher uncertainty (0.47–1.53) and dinoflagellates overall have the highest uncertainty (0.63–1.64). Regarding the two prokaryotic phytoplankton, the uncertainty for the direct retrieval of prokaryote Chl-a shows lower uncertainty in the polar regions but higher in the low latitudes (Figure 12g); the uncertainty for *Prochlorococcus* shows generally high uncertainty in latitudes higher than 35° both south and north and also in the gyres (Figure 12j). The uncertainties for the prokaryotic fractions (Figures 12h and 12k) show reverse distributions to the fraction retrievals with low uncertainty



**Figure 11.** Distributions of the actual error normalized by the modeled propagated error ( $\delta_{PFT}/\sigma_{PFT}$ ) for all the retrieved (using EOF-SST hybrid algorithm) PFT quantities from the testing data set: (a) TChl-a; Chl-a of (b) diatoms, (c) dinoflagellates, (d) haptophytes, (e) green algae, and (f) prokaryotes; (g) *f*-Prokaryotes, (h) fraction-derived prokaryote Chl-a, (i) *Prochlorococcus* Chl-a, (j) *f*-*Prochlorococcus*, and (k) fraction-derived *Prochlorococcus* Chl-a. Red and green curves indicate the fitted normal distribution and the standard centered normal distribution, respectively. The red asterisk represents the mean point of the fitted distribution, and the green circle highlights the point of zero. Mean value and standard deviation (SD) of  $\delta_y/\sigma_y$  are also shown. EOF, empirical orthogonal function; PFT, phytoplankton functional type; SST, sea surface temperature.

in midlatitude to low-latitude compared to that in high latitudes. The uncertainty for *f*-prokaryotes is lower than *f*-*Prochlorococcus* uncertainty. Fraction-derived prokaryote and *Prochlorococcus* Chl-a uncertainties basically follow the patterns of their corresponding fraction uncertainties but are higher (Figures 12i and 12j), as they are derived by combining both the fraction and TChl-a uncertainties, with the latter exhibiting little spatial variation (Figure 12a).



**Figure 12.** Per-pixel uncertainty (in natural logarithmic scale) of the annual mean of 2011 for the satellite-derived PFT quantities from EOF-SST hybrid algorithm. The dashed-line box frames in particular (a) the uncertainty of TChl-a from the EOF-SST hybrid algorithm in comparison to (d) the OC-CCI TChl-a uncertainty. EOF, empirical orthogonal function; PFT, phytoplankton functional type; SST, sea surface temperature.

As the PFT prediction models are based on multiple linear regressions, the uncertainty of the model coefficients and SST are propagated linearly to the PFT retrievals following Equation 8. Their corresponding uncertainties are found to have much less spatial variability compared to  $\sigma_{y(R_{rs})}$ . As a result, the distribution patterns of the pixel-wise uncertainties generated for the PFT quantities are very much subject to  $\sigma_{y(R_{rs})}$  derived from the linear LUT functions. It should be noted that the LUTs built for prokaryote and *Prochlorococcus* Chl-a cannot represent their uncertainties sufficiently (Figures 10f and 10g), hence their uncertainty products (Figures 12g and 12j) should be used with caution. The uncertainties for the fractions and the fraction-derived prokaryote and *Prochlorococcus* Chl-a, however, are reported with higher confidence as they are well described by the LUTs and the error propagation analysis. Uncertainties for the other PFT quantities are in general well and conservatively quantified, which are also justified by the uncertainty validation. Yet uncertainties for satellite-derived PFTs have been rarely reported, with the exception of Brewin, Ciavatta, et al. (2017), who performed uncertainty evaluation on phytoplankton size classes and two phytoplankton groups (diatoms and dinoflagellates) retrieved by the returned abundance/ecological based algorithm in the North Atlantic region. Though the exact values of the uncertainty estimates are not provided by Brewin, Ciavatta, et al. (2017), their uncertainty maps for diatoms and dinoflagellates are within the same order of magnitudes (when converted from the base-10 logarithm to the natural-logarithm) and show similar distribution patterns with our uncertainty products in the North Atlantic Ocean.

Compared to the reported uncertainty for the OC-CCI TChl-a product, the uncertainty of TChl-a from our EOF-SST hybrid algorithm is considerably lower. Our TChl-a uncertainty varies from 0.52 to 0.58 (Figure 12a), presenting very little spatial variability with only slightly higher uncertainty in the gyres and some marginal seas. The OC-CCI TChl-a uncertainty ranges between 0.43 and 1.18 (Figure 12d), showing large spatial variability with much higher uncertainty in the marginal seas and high latitudes compared to our TChl-a uncertainty. Low uncertainty for the OC-CCI product in and surrounding the gyres is comparable with our uncertainty estimates. Remarkably reduced uncertainty of the TChl-a derived by the EOF-SST hybrid algorithm in high latitudes ( $>40^\circ$ ) indicates that the EOF-SST hybrid algorithm has great potential in improving TChl-a estimation, especially in polar regions where the standard OC algorithms typically introduce high errors (IOCCG, 2015). The overall lower uncertainty in our TChl-a product also reveals at a certain scale that our estimates for most of the PFT quantities and the corresponding uncertainties are reliable.

### 3.3.4. Potential Caveat in the Uncertainty Analysis

As mentioned in Section 2.3, we did not include the uncertainties of the HPLC measurements and the DPA-derived PFTs in our PFT per-pixel uncertainty construction. According to IOCCG (2019), uncertainties of the HPLC-based Chl-a are around 7% and can be higher for other pigments (Claustre et al., 2004). We were not able to obtain this information for our large global data set collected from various cruises in the last decades. Mostly no other descriptors of phytoplankton taxonomic composition had been measured and details on the HPLC measurement error (including all associated steps, e.g., filtration, extraction, and HPLC analysis accuracy) are not available. Nonetheless, HPLC-based pigment measurements still remain as one of the most widespread and quality-controlled methods currently available, and are robust for quantifying pigments and their concentrations (Van Heukelem & Hooker, 2011). However, the methods that transform the pigment concentrations into PFT quantities rely on assumptions that might be violated (Moore & Brown, 2020). There are still large uncertainties in how we characterize the PSCs and PFTs in situ (e.g., Brewin et al., 2014; Chase et al., 2020). We used HPLC data to derive PFT Chl-a and fractions by determining the DPs, for which the DPA method was updated by Losa et al. (2017) based on previous refinements (e.g., Brewin et al., 2015; Hirata et al., 2011; Uitz et al., 2006). It is known that marker pigments used for PFT identification are often shared among taxonomic groups, bringing uncertainty to the quantification of PFTs (Jeffrey et al., 2011). A recent study by Chase et al. (2020) examined the PSCs derived from DPA by comparing to the cytometry data and found that DPA tended to mis-assign several major phytoplankton groups to size classes. They provided recommendations to revise these mis-assignments by adjusting the attribution of phytoplankton groups to different PSCs, also Hirata et al. (2011) and Brewin et al. (2014) have pointed to that. As we directly used the DPs for our PFT determination and did not assign the DPs to size classes (see Section 2.3.1.1); unfortunately, these recommendations by Chase et al. (2020) and other researchers do not apply to our study. However, validations of DPA-based PFT derivations using data from other sources such as microscopy and flow cytometry are always desirable, although we still have to bear in mind that different quantities are determined and also their conversion bear large uncertainties (Bracher et al., 2017).

Using both global and regional HPLC data sets, Kramer and Siegel (2019) performed hierarchical cluster and EOF analyses to examine the associates between and among pigment groups, and demonstrated that four major phytoplankton groups (diatoms/dinoflagellates, prokaryotes, haptophytes, and green algae) could be identified on global scales. Their results showed that diatoms and dinoflagellates grouped together on global scales but could always be separated on local scales. Our data set includes many if not most of the data sets of Kramer & Siegel (2019) both from global and local regions, which was used to determine six PFTs. Therefore, according to their findings, our dinoflagellate Chl-a (using peridinin as DP) and *Prochlorococcus* Chl-a (using divinyl chlorophyll *a* as DP) quantifications might not be confident. Interestingly, this is reflected in our PFT uncertainty maps which show that retrievals of dinoflagellate and *Prochlorococcus* Chl-a have relatively higher uncertainty ranges than others (Figures 12c and 12l), whereas haptophytes and green algae show the lowest uncertainty ranges (Figures 12e and 12f). These uncertainty maps, though without directly taking into account the errors in the in situ PFTs due to DPA, correspond to the results of assessment performed by Kramer & Siegel (2019) and thus carry partly the uncertainty due to DPA through the retrieval algorithms. Nevertheless, it is still necessary to reconstruct the uncertainty assessment with the attribution of the in situ PFT uncertainties when they are better understood.

#### 4. Summary

This study improved the previously established EOF-based statistical approach by Xi et al. (2020) for globally estimating the Chl-a of six PFTs using merged OC  $R_{rs}$  products. The modified retrieval scheme, referred to as EOF-SST hybrid algorithm, was developed by using updated input data sets and accounting for the influence of SST on the derived PFT quantities. Furthermore, fractions of prokaryotes and *Prochlorococcus* were also included as retrieved PFT quantities, which resulted in more accurate retrievals compared to retrieving their Chl-a directly. The retrieved fractions were used together with TChl-a retrievals to obtain a fraction-derived Chl-a for the two PFTs, which showed prominent improvements for prokaryote Chl-a, but not for *Prochlorococcus* Chl-a. By further splitting the input data set according to the PFT dependence on SST in different SST regimes, separated retrieval algorithms for low and high-temperature waters were established, leading to further improved performance for all PFTs than the hybrid algorithm based on the whole data set. Improvements for PFT retrievals were largest in high latitudes. Finally, the pixel-by-pixel uncertainty of the satellite PFT retrievals was assessed by accounting for the uncertainties from input data and model parameters via an error propagation method. These satellite PFT uncertainties, for the first time reported on a global scale and for spectral-based PFT retrieval approaches, provide reliable error estimates for the PFT products which allow us to better understand the product quality both in time and space.

This study used the GlobColour merged OC products spanning a period from 2002 to 2012 only. However, our EOF-SST hybrid algorithm including pixelwise uncertainties can easily be expanded to other OC sensors, such as MODIS-VIIRS merged and OLCI products, encompassing information at least at nine spectral bands. The algorithm can also be adapted and well-suited to hyperspectral OC data (e.g., PACE); its performance for PFT retrievals may potentially be improved with higher spectral resolution and incorporation of environmental information. Uncertainty assessment for the PFT estimates from different satellite products is needed for consistent long-term PFT data from multiple satellite sensors with assured continuity. Such data are critical to track changes in phytoplankton community structure under the changing climate. PFT products with uncertainty estimates are also beneficial to applications of ecosystem modeling by helping to simulate and/or evaluate the model outputs, as well as being assimilated within these models to further improve forecasting marine biogeochemistry, as used by many marine services.

#### Data Availability Statement

The DPA-derived PFT Chl-a for diatoms, haptophytes, and prokaryotes from the pigment database I were published already in Losa et al. (2017) and are available from PANGAEA: <https://doi.pangaea.de/10.1594/PANGAEA.875879> (Soppa et al., 2017). Table S1 lists the data sources where the global HPLC data were obtained. Matchup data used for the algorithm establishment in this study, including the collocated in situ pigment-derived PFT data, CMEMS SST, and nine-band  $R_{rs}$  data from the GlobColour merged products, are available from PANGAEA: <https://doi.pangaea.de/10.1594/PANGAEA.930087> (Xi et al., 2021).

#### References

- ACRI-ST GlobColour Team, Mangin, A., & Hembise Fanton d'Andon, O. (2017). *GlobColour product user Guide* (GC-UM-ACR-PUG-01, version 4.1). ACRI-ST
- Aiken, J., Pradhan, Y., Barlow, R., Lavender, S., Poulton, A., Holligan, P., & Hardman-Mountford, N. (2009). Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995–2005. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(15), 899–917. <https://doi.org/10.1016/j.dsr2.2008.09.017>
- Alvain, S., Moulin, C., Dandonneau, Y., & Bréon, F. M. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep Sea Research Part I: Oceanographic Research Papers*, 52(11), 1989–2004. <https://doi.org/10.1016/j.dsr.2005.06.015>
- Alvain, S., Moulin, C., Dandonneau, Y., & Loisel, H. (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. *Global Biogeochemical Cycles*, 22(3). <https://doi.org/10.1029/2007GB003154>
- Amante, C., & Eakins, B. W. (2009). *ETOPO1 1 arc-minute global relief model: Procedures, data sources and analysis* (NOAA technical memorandum NESDIS NGDC-24). National Geophysical Data Center.
- Antoine, D., d'Ortenzio, F., Hooker, S. B., Bécu, G., Gentili, B., Tailliez, D., & Scott, A. J. (2008). Assessment of uncertainty in the ocean reflectance determined by three satellite ocean color sensors (MERIS, SeaWiFS and MODIS-A) at an offshore site in the Mediterranean Sea (BOUSSOLE project). *Journal of Geophysical Research*, 113, C07013. <https://doi.org/10.1029/2007JC004472>
- Bracher, A., Bouman, H. A., Brevin, R. J. W., Bricaud, A., Brotas, V., Ciotti, A. M., et al. (2017). Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development. *Frontiers in Marine Science*, 4, 1–15. <https://doi.org/10.3389/fmars.2017.00055>

#### Acknowledgments

This study is supported by a collaborative project, OLCI-PFT (ACRI-AWI Offer #209-180104), between ACRI-ST and Phytooptics team at Alfred Wegener Institute. The contribution by Svetlana N. Losa was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project number 268020496—TRR 172, within the Transregional Collaborative Research Center ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)3 (Project C03), and was also partially made in the framework of the state assignment of the Federal Agency for Scientific Organizations (FASO) Russia (theme 0128-2021-0014). The authors thank Marc Taylor for the primary R script on the EOF method, Sonja Wiegmann for her dedication in measuring and processing the HPLC data for our own cruises, and all the previous and current Phytooptics team members who participated in the past cruises for data collection and analysis. The authors are thankful to all the scientists and crew who were involved in the global HPLC pigment data collection and analyses for making their data publicly available (see Table S1). The authors also thank NASA, ESA, and EUMETSAT for the SeaWiFS, MODIS, and MERIS data, and especially the GlobColour program for providing the merged ocean color L3 products. GlobColour data (<http://globcolour.info>) used in this study has been developed, validated, and distributed by ACRI-ST, France. This study has been conducted using E.U. Copernicus Marine Service Information in terms of CMEMS sea surface temperature products. The authors also thank the Ocean-Color Climate Change Initiative (OC-CCI) program for providing global chlorophyll-a data products. The authors finally acknowledge the three anonymous reviewers who helped to improve this study by providing valuable and constructive comments. Open access funding enabled and organized by Projekt DEAL.

- Bracher, A., Taylor, M. H., Taylor, B., Dinter, T., Röttgers, R., & Steinmetz, F. (2015). Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Ocean Science*, *11*, 139–158. <https://doi.org/10.5194/os-11-139-2015>
- Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., & Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences*, *6*, 751–764. <https://doi.org/10.5194/bg-6-751-2009>
- Bracher, A., Xi, H., Dinter, T., Mangin, A., Strass, V. H., von Appen, W.-J., & Wiegmann, S. (2020). High resolution water column phytoplankton composition across the Atlantic Ocean from ship-towed vertical undulating radiometry. *Frontiers in Marine Science*, *7*, 235. <https://doi.org/10.3389/fmars.2020.00235>
- Brewin, R. J. W., Ciavatta, S., Sathyendranath, S., Jackson, T., Tilstone, G., Curran, K., et al. (2017). Uncertainty in ocean-color estimates of chlorophyll for phytoplankton groups. *Frontiers in Marine Science*, *4*, 104. <https://doi.org/10.3389/fmars.2017.00104>
- 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. <https://doi.org/10.1016/j.ecolmodel.2010.02.014>
- 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. <https://doi.org/10.1016/j.rse.2015.07.004>
- 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. <https://doi.org/10.1016/j.dsr.2013.11.007>
- Brewin, R. J. W., Tilstone, G. H., Jackson, T., Cain, T., Miller, P. I., Lange, P. K., et al. (2017). Modelling size-fractionated primary production in the Atlantic Ocean from remote sensing. *Progress in Oceanography*, *158*, 130–149. <https://doi.org/10.1016/j.pocean.2017.02.002>
- Cael, B. B., Chase, A., & Boss, E. (2020). Information content of absorption spectra and implications for ocean color inversion. *Applied Optics*, *59*(13), 3971–3984. <https://doi.org/10.1364/AO.3891893>
- Chase, A. P., Kramer, S. J., Haëntjens, N., Boss, E. S., Karp-Boss, L., Edmondson, M., & Graff, J. R. (2020). Evaluation of diagnostic pigments to estimate phytoplankton size classes. *Limnology and Oceanography: Methods*, *18*(10), 570–584. <https://doi.org/10.1002/lom3.10385>
- Claustre, H., Hooker, S. B., Van Heukelem, L., Berthon, J.-F., Barlow, R., Ras, J., et al. (2004). An intercomparison of HPLC phytoplankton pigment methods using in situ samples: Application to remote sensing and database activities. *Marine Chemistry*, *85*(1–2), 41–61. <https://doi.org/10.1016/j.marchem.2003.09.002>
- Correa-Ramirez, M., Morales, C., Letelier, R., Anabalón, V., & Hormazabal, S. (2018). Improving the remote sensing retrieval of phytoplankton functional types (PFT) using empirical orthogonal functions: A case study in a coastal upwelling region. *Remote Sensing*, *10*(4), 498. <https://doi.org/10.3390/rs10040498>
- Craig, S. E., Jones, C. T., Li, W. K. W., Lazin, G., Horne, E., Caverhill, C., & Cullen, J. J. (2012). Deriving optical metrics of coastal phytoplankton biomass from ocean colour. *Remote Sensing of Environment*, *119*, 72–83. <https://doi.org/10.1016/j.rse.2011.12.007>
- de Mora, L., Butenschön, M., & Allen, J. I. (2016). The assessment of a global marine ecosystem model on the basis of emergent properties and ecosystem function: A case study with ERSEM. *Geoscientific Model Development*, *9*, 59–76. <https://doi.org/10.5194/gmd-9-59-2016>
- Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, *116*, 140–158. <https://doi.org/10.1016/j.rse.2010.10.017>
- Dutkiewicz, S., Morris, J. J., Follows, M. J., Scott, J., Levitan, O., Dyhrman, S. T., & Berman-Frank, I. (2015). Impact of ocean acidification on the structure of future phytoplankton communities. *Nature Climate Change*, *5*, 1002–1006. <https://doi.org/10.1038/nclimate2722>
- Falkowski, P. G., Laws, E. A., Barber, R. T., & Murray, J. W. (2003). Phytoplankton and their role in primary, new, and export production. In M. J. R. Fasham (Ed.), Ed., *Ocean biogeochemistry. Global change—The IGBP series (closed)* (pp. 99–121). Springer. [https://doi.org/10.1007/978-3-642-55844-3\\_5](https://doi.org/10.1007/978-3-642-55844-3_5)
- Flombaum, P., Gallegos, J. L., Gordillo, R. A., Rincon, J., Zabala, L. L., Jiao, N., et al. (2013). Present and future global distributions of the marine Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 9824–9829. <https://doi.org/10.1073/pnas.1307701110>
- Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., et al. (2011). Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, *8*, 311–327. <https://doi.org/10.5194/bg-8-311-2011>
- Hirata, T., Saux-Picart, S., Hashioka, T., Aita-Noguchi, M., Sumata, H., Shigemitsu, M., et al. (2013). A comparison between phytoplankton community structures derived from a global 3D ecosystem model and satellite observation. *Journal of Marine Systems*, *109*–*110*, 129–137. <https://doi.org/10.1016/j.jmarsys.2012.01.009>
- Holt, J., Icarus Allen, J., Anderson, T. R., Brewin, R., Butenschön, M., Harle, J., et al. (2014). Challenges in integrative approaches to modelling the marine ecosystems of the North Atlantic: Physics to fish and coasts to ocean. *Progress in Oceanography*, *129*, 285–313. <https://doi.org/10.1016/j.pocean.2014.04.024>
- IOCCG (2014). Phytoplankton functional types from space. In S. Sathyendranath, & V. Stuart (Eds.), Eds., *Reports of the international ocean color coordinating group* (No. 15). IOCCG.
- IOCCG (2015). Ocean color remote sensing in polar seas. In M. Babin, K. Arriago, S. Bélanger, & M.-H. Forget (Eds.), Eds., *Reports of the international ocean color coordinating group* (No. 16). IOCCG.
- IOCCG (2019). Uncertainties in ocean color remote sensing. In F. Melin (Ed.), Ed., *Reports of the international ocean color coordinating group* (No. 18). IOCCG.
- JCGM (2008). Evaluation of measurement data – Guide to the expression of uncertainty in measurement. *Joint Committee for Guides in Metrology*, *100*, 2008.
- Jeffrey, S. W., Wright, S. W., & Zapata, M. (2011). Microalgal classes and their signature pigments. In S. Roy, C. A. Llewellyn, E. S. Egeland, & G. Johnsen (Eds.), Eds., *Phytoplankton pigments: Characterization, chemotaxonomy, and application in oceanography* (pp. 3–77). Cambridge University Press.
- Johnson, R., Strutton, P. G., Wright, S. W., McMinn, A., & Meiners, K. M. (2013). Three improved satellite chlorophyll algorithms for the Southern Ocean. *Journal of Geophysical Research: Oceans*, *118*, 3694–3703. <https://doi.org/10.1002/jgrc.20270>
- Kostadinov, T. S., Milutinović, S., Marinov, I., & Cabré, A. (2016). Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution. *Ocean Science*, *12*, 561–575. <https://doi.org/10.5194/os-12-561-2016>
- 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? *Journal of Geophysical Research: Oceans*, *124*, 7557–7574. <https://doi.org/10.1029/2019JC015604>

- Lange, P. K., Werdell, P. J., Erickson, Z. K., Dall'Omo, G., Brewin, R. J. W., Zubkov, M. V., et al. (2020). Radiometric approach for the detection of picophytoplankton assemblages across oceanic fronts. *Optics Express*, *28*(18), 25682–25705. <https://doi.org/10.1364/OE.398127>
- Le Quéré, C., Harrison, S. P., Prentice, C. I., Buitenhuis, E. T., Aumonts, O., Bopp, L., et al. (2005). Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology*, *11*, 2016–2040. <https://doi.org/10.1111/j.1365-2486.2005.1004.x>
- Lee, Z., Arnone, R., Hu, C., Werdell, P. J., & Lubac, B. (2010). Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Applied Optics*, *49*, 369–381. <https://doi.org/10.1364/AO.49.000369>
- Losa, S. N., Dutkiewicz, S., Losch, M., Oelker, J., Soppa, M. A., Trimborn, S., et al. (2019). On modeling the Southern Ocean phytoplankton functional types. *Biogeosciences Discussions*. <https://doi.org/10.5194/bg-2019-289>
- Losa, S. N., Soppa, M. A., Dinter, T., Wolanin, A., Brewin, R. J. W., Bricaud, A., et al. (2017). Synergistic Exploitation of Hyper- and Multi-Spectral Precursor Sentinel Measurements to Determine Phytoplankton Functional Types (SynSenPFT). *Frontiers in Marine Science*, *4*, 1–22. <https://doi.org/10.3389/fmars.2017.00203>
- Lubac, B., & Loisel, H. (2007). Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Remote Sensing of Environment*, *110*(1), 45–58. <https://doi.org/10.1016/j.rse.2007.02.012>
- Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, *114*, 1791–1804. <https://doi.org/10.1016/j.rse.2010.04.002>
- McKinna, L. I. W., Cetinic, I., Chase, A. P., & Werdell, P. J. (2019). Approach for propagating radiometric data uncertainties through NASA ocean color algorithms. *Frontiers of Earth Science*, *7*, 176. <https://doi.org/10.3389/feart.2019.00176>
- Mélin, F., & Franz, B. A. (2014). Assessment of satellite ocean colour radiometry and derived geophysical products. In G. Zibordi, C. Donlon, & A. Parr (Eds.), *Optical radiometry for oceans climate measurements, experimental methods in the physical sciences* (Vol. 47). Academic Press.
- Moore, T. S., & Brown, C. W. (2020). Incorporating environmental data in abundance-based algorithms for deriving phytoplankton size classes in the Atlantic Ocean. *Remote Sensing of Environment*, *240*, 111689. <https://doi.org/10.1016/j.rse.2020.111689>
- 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*, 540–558. <https://doi.org/10.1029/2018GB006118>
- Mouw, C. B., Hardman-Montford, N., Alvain, S., Bracher, A., Brewin, R. J. W., Bricaud, A., et al. (2017). A consumer's guide to satellite remote sensing of multiple phytoplankton groups in the global ocean. *Frontiers in Marine Science*, *4*, 00041. <https://doi.org/10.3389/fmars.2017.00041>
- Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., & Nerger, L. (2019). Assimilation of global total chlorophyll OC-CCI data and its impact on individual phytoplankton fields. *Journal of Geophysical Research: Oceans*, *124*, 470–490. <https://doi.org/10.1029/2018JC014329>
- Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., & Nerger, L. (2020). Global assimilation of ocean-color data of phytoplankton functional types: Impact of different datasets. *Journal of Geophysical Research: Oceans*, *125*.e2019JC015586. <https://doi.org/10.1029/2019JC015586>
- Qi, L., Lee, Z., Hu, C., & Wang, M. (2017). Requirement of minimal signal-to-noise ratios of ocean color sensors and uncertainties of ocean color products. *Journal of Geophysical Research: Oceans*, *122*, 2595–2611. <https://doi.org/10.1002/2016JC012558>
- Sathyendranath, S., Brewin, R., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., et al. (2019). An ocean-colour time series for use in climate studies: The experience of the Ocean-Colour Climate Change Initiative (OC-CCI). *Sensors*, *19*, 4285. <https://doi.org/10.3390/s19194285>
- Sathyendranath, S., Jackson, T., Brockmann, C., Brotas, V., Calton, B., Chuprin, A., et al. (2020). *ESA ocean color climate change initiative (Ocean\_Colour\_cci): Global chlorophyll-a data products gridded on a sinusoidal projection*. Centre for Environmental Data Analysis. Retrieved from <https://catalogue.ceda.ac.uk/uuid/99348189bd33459cbd597a58c30d8d10>
- Soja-Woźniak, M., Craig, S. E., Kratzer, S., Wojtasiewicz, B., Darecki, M., & Jones, C. T. (2017). A novel statistical approach for ocean color estimation of inherent optical properties and cyanobacteria abundance in optically complex waters. *Remote Sensing*, *9*, 1–22. <https://doi.org/10.3390/rs9040343>
- Soppa, M. A., Hirata, T., Silva, B., Dinter, T., Peeken, I., Wiegmann, S., & Bracher, A. (2014). Global retrieval of diatom abundance based on phytoplankton pigments and satellite data. *Remote Sensing*, *6*, 10089–10106. <https://doi.org/10.3390/rs61010089>
- Soppa, M. A., Peeken, I., & Bracher, A. (2017). Global chlorophyll “a” concentrations for diatoms, haptophytes and prokaryotes obtained with the diagnostic pigment analysis of HPLC data compiled from several databases and individual cruises. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.875879>
- Taylor, B. B., Taylor, M. H., Dinter, T., & Bracher, A. (2013). Estimation of relative phycoerythrin concentrations from hyperspectral underwater radiance measurements—A statistical approach. *Journal of Geophysical Research: Oceans*, *118*, 2948–2960. <https://doi.org/10.1002/jgrc.20201>
- 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*, C08005. <https://doi.org/10.1029/2005JC003207>
- Van Heukelem, L., & Hooker, S. B. (2011). The importance of a quality assurance plan for method validation and minimizing uncertainties in the HPLC analysis of phytoplankton pigments. In S. Roy, C. A. Llewellyn, E. S. Egeland, & G. Johnsen (Eds.), *Phytoplankton pigments: Characterization, chemotaxonomy, and applications in oceanography* (pp. 195–242). Cambridge University Press.
- Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., & Marty, J.-C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research: Oceans*, *106*(C9). <https://doi.org/10.1029/1999JC000308>
- Ward, B. A. (2015). Temperature-correlated changes in phytoplankton community structure are restricted to polar waters. *PLOS One*, *10*(8), e0135581. <https://doi.org/10.1371/journal.pone.0135581>
- Ward, B. A., Dutkiewicz, S., Jahn, O., & Follows, M. J. (2012). A size-structured food-web model for the global ocean. *Limnology & Oceanography*, *57*, 1877–1891. <https://doi.org/10.4319/lo.2012.57.6.1877>
- Werdell, P. J., McKinna, L. I. W., Boss, E., Ackleson, S. G., Craig, S. E., Gregg, W. W., et al. (2018). An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. *Progress in Oceanography*, *160*, 186–212. <https://doi.org/10.1016/j.pocean.2018.01.001>
- Werdell, P. J., Roesler, C. S., & Goes, J. I. (2014). Discrimination of phytoplankton functional groups using an ocean reflectance inversion model. *Applied Optics*, *53*, 4833. <https://doi.org/10.1364/ao.53.004833>
- Worsfold, M., Good, S., McLaren, A., Fiedler, E., Roberts-Jones, J., & Martin, M. (2020). Quality information document global ocean OSTIA sea surface temperature reprocessing SST-GLO-SST-L4-REP-OBSERVATIONS-010-011. Ref: CMEMS-SST-QUID-010-011.

- Xi, H., Losa, S. N., Mangin, A., Garnesson, P., Bretagnon, M., Demaria, J., et al. (2021). A data set of collocated satellite remote sensing reflectance from GlobColour merged products, chlorophyll a concentration of phytoplankton functional types derived from in situ pigment data, and CMEMS sea surface temperature from 2002 to 2012. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.930087>
- Xi, H., Losa, S. N., Mangin, A., Soppa, M. A., Garnesson, P., Demaria, J., et al. (2020). Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. *Remote Sensing of Environment*, *240*, 111704. <https://doi.org/10.1016/j.rse.2020.111704>
- Xiao, Y., & Friedrichs, M. A. M. (2014). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, *119*, 2691–2712. <https://doi.org/10.1002/2013jc009433>. <https://doi.org/10.1093/plankt/fbp127>

## References From the Supporting Information

- Brown, M. S., Munro, D. R., Feehan, C. J., Sweeney, C., Ducklow, H. W., & Schofield, O. M. (2019). Enhanced oceanic CO<sub>2</sub> uptake along the rapidly changing West Antarctic Peninsula. *Nature Climate Change*, *9*, 678–683. <https://doi.org/10.1038/s41558-019-0552-3>
- Clayton, S., Nagai, T., & Follows, M. J. (2014). Hydrochemistry and phytoplankton pigments in water samples obtained during R/V Natsushima (JAMSTEC) cruise to the Kuroshio Extension Front in October 2009. *PANGAEA*. Supplement 1 <https://doi.org/10.1594/PANGAEA.819108>
- Hepach, H., Quack, B., Tegtmeier, S., Engel, A., Bracher, A., Fuhlbrügge, S., et al. (2016). Pigment measured on water bottle samples during METEOR cruise M91. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.864786>
- Kozłowski, W. A., Deutschman, D., Garibotti, I., Trees, C., & Vernet, M. (2011). An evaluation of the application of CHEMTAX to Antarctic coastal pigment data. *Deep-Sea Research I*, *58*, 350–364. <https://doi.org/10.1016/j.dsr.2011.01.008>
- Schofield, O., Saba, G., Coleman, K., Carvalho, F., Couto, N., Ducklow, H., et al. (2017). Decadal variability in coastal phytoplankton community composition in a changing West Antarctic Peninsula. *Deep-Sea Research Part I Oceanographic Research Papers*, *124*, 42–54. <https://doi.org/10.1016/j.dsr.2017.04.014>
- Vincent, W. F. (2000). Cyanobacterial dominance in the polar regions. In B. A. Whitton, & M. Potts (Eds.), Eds., *The ecology of cyanobacteria*. Springer. [https://doi.org/10.1007/0-306-46855-7\\_12](https://doi.org/10.1007/0-306-46855-7_12)
- Wilkins, D., Yau, S., Williams, T. J., Allen, M. A., Brown, M. V., DeMaere, M. Z., et al. (2013). Key microbial drivers in Antarctic aquatic environments. *FEMS Microbiology Reviews*, *37*(3), 303–335. <https://doi.org/10.1111/1574-6976.12007>