remote sensing

MDPI

Article

Retrieval of Phytoplankton Pigment Composition from Their In Vivo Absorption Spectra

Yinxue Zhang [1,2], Guifen Wang [1,2,*], Shubha Sathyendranath [3], Wenlong Xu [1,2], Yizhe Xiao [1,2] and Long Jiang [1,2]

1 Key Laboratory of Marine Hazards Forecasting, Ministry of Natural Resources, Hohai University, Nanjing 210098, China; zhangyinxue@hhu.edu.cn (Y.Z.); wenlongxu@hhu.edu.cn (W.X.); yizhexiao@hhu.edu.cn (Y.X.); ljiang@hhu.edu.cn (L.J.)
2 College of Oceanography, Hohai University, Nanjing 210098, China
3 National Centre for Earth Observation, Plymouth Marine Laboratory, Prospect Place, Plymouth PL1 3DH, UK; ssat@pml.ac.uk
* Correspondence: guifenwang@hhu.edu.cn; Tel.: +86-135-4438-8059

Abstract: Algal pigment composition is an indicator of phytoplankton community structure that can be estimated from optical observations. Assessing the potential capability to retrieve different types of pigments from phytoplankton absorption is critical for further applications. This study investigated the performance of three models and the utility of hyperspectral in vivo phytoplankton absorption spectra for retrieving pigment composition using a large database (n = 1392). Models based on chlorophyll-a (Chl-a model), Gaussian decomposition (Gaussian model), and partial least squares (PLS) regression (PLS model) were compared. Both the Gaussian model and the PLS model were applied to hyperspectral phytoplankton absorption data. Statistical analysis revealed the advantages and limitations of each model. The Chl-a model performed well for chlorophyll-c (Chl-c), diadinoxanthin, fucoxanthin, photosynthetic carotenoids (PSC), and photoprotective carotenoids (PPC), with a median absolute percent difference for cross-validation ($MAPD_{CV}$) < 58%. The Gaussian model yielded good results for predicting Chl-a, Chl-c, PSC, and PPC ($MAPD_{CV}$ < 43%). The performance of the PLS model was comparable to that of the Chl-a model, and it exhibited improved retrievals of chlorophyll-b, alloxanthin, peridinin, and zeaxanthin. Additional work undertaken with the PLS model revealed the prospects of hyperspectral-resolution data and spectral derivative analyses for retrieving marker pigment concentrations. This study demonstrated the applicability of in situ hyperspectral phytoplankton absorption data for retrieving pigment composition and provided useful insights regarding the development of bio-optical algorithms from hyperspectral and satellite-based ocean-colour observations.

Keywords: hyperspectral phytoplankton absorption; multispectral phytoplankton absorption; phytoplankton pigment concentration; derivative spectra; bio-optical algorithms

check for updates

Citation: Zhang, Y.; Wang, G.; Sathyendranath, S.; Xu, W.; Xiao, Y.; Jiang, L. Retrieval of Phytoplankton Pigment Composition from Their In Vivo Absorption Spectra. *Remote Sens.* **2021**, *13*, 5112. https://doi.org/10.3390/rs13245112

Academic Editor: Malgorzata Stramska

Received: 9 November 2021
Accepted: 13 December 2021
Published: 16 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The magnitude of oceanic carbon fixation by phytoplankton through photosynthesis is comparable to net production by terrestrial plants at the global scale, making phytoplankton a key component in the global carbon cycle. Phytoplankton dynamics are important when examining existential threats such as the greenhouse effect and ocean acidification [1–4]. The role of phytoplankton in various biogeochemical cycles varies with their functional type. Important phytoplankton types can be distinguished through auxiliary pigments that are referred to as marker pigments. It is a basic requirement in the field of marine ecology to study the quantitative distribution of these marker pigments at large spatiotemporal scales. Their concentrations, together with that of the main photosynthetic pigment, chlorophyll-a (Chl-a), can be estimated using high-performance liquid chromatography (HPLC) techniques. However, HPLC methods are time-consuming and expensive, limiting

the number of observations that can be made. During the last few decades, optical data, including reflectance and spectral absorption measurements from various platforms have been successfully used to determine the concentration of Chl-a as well as other pigments at large spatiotemporal scales [3,5–9]. Many approaches have been developed to retrieve pigment compositions of phytoplankton from multispectral or hyperspectral data.

Most current ocean-colour sensors provide spectral information on remote sensing reflectance (R_{rs}) in the visible domain at a limited set of wavebands. However, the potential ability of hyperspectral retrieval of phytoplankton pigment compositions has attracted much attention [6,10], especially with the launch of the PRISMA satellite [11,12] and the upcoming PACE missions of NASA [13]. Algorithms using empirical orthogonal functions and Gaussian decomposition have been designed to improve retrieval of phytoplankton pigments from in situ hyperspectral R_{rs} data [6,7,10,14], preparatory to their use in hyperspectral satellite missions. Their results showed the potential utility of optical methods for retrieval of certain types of pigments and suggested that using continuous hyperspectral instruments was beneficial for improving algorithm performance. However, challenges of finding a global set of have also been raised. With an inversion algorithm that defines phytoplankton absorption spectra ($a_{\mathrm{B}}(\lambda)$, where we have used the subscript B to indicate phytoplankton biomass) as a sum of Gaussian functions, Chase et al. [7] found that this method is only feasible for a limited set of pigments (Chl-a; chlorophyll-b, Chl-b; chlorophyll-c_1 + chlorophyll-c_2, Chl-$c_{1,2}$; and photoprotective carotenoids, PPC) and similar errors were found for empirical algorithms with Chl-a as an intermediate variable. Xi et al. [15] re-tuned the empirical orthogonal function algorithm by including sea surface temperature as an adaptational input parameter.

Algorithms for retrieval of phytoplankton pigments from R_{rs} typically rely on the influence of pigments on the spectral absorption characteristics of phytoplankton [6,7,16–18], or empirical relationships between the concentration of Chl-a and the major auxiliary pigments [19]. Phytoplankton-absorption-based methods make possible a more direct approach to predict phytoplankton community structures. Many studies focused on phytoplankton absorption spectra to investigate the possibilities of using hyperspectral data to get more information about pigment composition [5,8,20]. Decomposition of $a_{\mathrm{B}}(\lambda)$ into Gaussian components attributed to specific pigments or pigment groups was first proposed by Hoepffner and Sathyendranath [21] and was successfully extended for partitioning particulate absorption ($a_{\mathrm{p}}(\lambda)$) data measured with underway flow-through system [5,9] and for characterizing the specific absorption spectra of eight phytoplanktonic groups [20]. Derivative analysis was used as another promising approach for isolating phytoplankton absorption signatures from hyperspectral ocean colour observations [14,22]. Organelli et al. [23] used the multivariate least squares regression technique to estimate phytoplankton size classes from the fourth-derivative spectra of phytoplankton or total particulate absorption. Catlett et al. [8] developed a novel optical model with principal components regression based on the unique relationships between pigments and absorption signatures isolated using spectral derivative analysis of $a_{\mathrm{B}}(\lambda)$. The method provided robust results for obtaining concentrations of many specific pigments. In short, these algorithms all have benefits and limitations. Understanding the advantages and limitations of different models may help us to assess the relative merits of various methods of pigment retrieval from optical data.

In this study, based on a large database of phytoplankton absorption and pigment composition data collected in three different areas (Northwest Atlantic, the Arabian Sea, and off the coast of Chile), we compare the performance of three model types for retrieval of pigment concentrations from in vivo $a_{\mathrm{B}}(\lambda)$ or from the concentration of Chl-a. The three types of retrieval models investigated here comprised (i) Chl-a models (of the type proposed by Hirata et al. [19] and Chase et al. [7]) that rely on observed relationships between the Chl-a concentration and the concentration of various auxiliary pigments; (ii) the Gaussian decomposition model (Gaussian model) [5,21,24,25], in which absorption spectra are broken down into a number of Gaussian components and their relationships with pigments

explored; and (iii) the partial least squares model (PLS model), in which concentrations of various pigments are expressed as linear functions of phytoplankton absorption at a number of key wavelengths [23], which have first been subjected to principal component analysis. The advantages of derivative analysis and the superiority of hyperspectral data over multispectral are also discussed.

2. Materials and Methods

2.1. Datasets

From 1994 to 2003, 3205 samples of phytoplankton pigment and absorption data were collected during 27 cruises. Sampling sites were located in the Northwest Atlantic, the Arabian Sea, and off the coast of Chile with water types ranging from eutrophic coastal waters to stratified and oligotrophic waters (Figure 1). Water samples were collected using Niskin bottles from the surface to a maximum depth of 140 m. In this study, 85% of the samples considered were collected within the top 50-m layer. Additional information regarding the measurement methods and the cruise's information is available in Sathyendranath et al. [26].

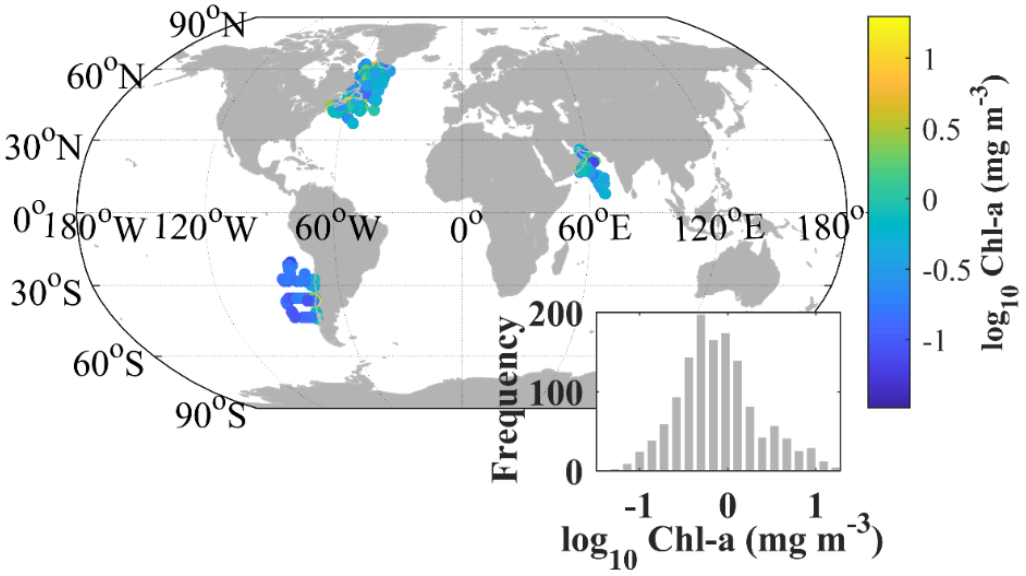


Figure 1. Locations of sites where phytoplankton absorption spectra were measured together with concentrations of major pigments. Inset shows the distribution of all log-transformed chlorophyll-a (Chl-a) data.

2.1.1. Phytoplankton Absorption Spectra

Particulate samples were collected on GF/F glass fiber filters and stored at −70 °C until processed. The quantitative filter-pad method was used for measuring $a_p(\lambda)$ over a range of wavelengths from 350 to 750 nm (at 1-nm increments), with a spectrophotometer equipped with a 15-cm integrating sphere. Phytoplankton pigments were extracted from the filters using a mixture of 90% acetone and dimethyl sulfoxide (6:4 vol:vol) for some cruises [27,28] or using 20 mL hot methanol for other cruises [26]. An exponential curve was fitted to the absorption spectrum of the residual material on the filter, to obtain the detrital absorption spectrum ($a_d(\lambda)$) [27,29]. The detrital absorption was subtracted from the particulate absorption spectrum to obtain $a_B(\lambda)$. This method of calculating phytoplankton absorption spectra from total particulate absorption spectra corrects for the effect of incomplete extraction of pigments on the absorption spectrum of the residual material. Correction for path-length amplification was implemented using the method of Hoepffner and Sathyendranath [24,30], with modifications as in Kyewalyanga et al. [31].

2.1.2. Phytoplankton Pigments

Pigment composition was analyzed using reverse-phase HPLC, as described in Head and Horne [32]. Filters were homogenized in 1.5 mL of 90% acetone, then centrifuged and diluted with 0.5 M ammonium acetate buffer at a ratio of 2:1 before injection. The samples were run on a Beckman C18 reverse-phase, 3-um Ultrasphere column (70 × 4.6 mm), using methanol/0.5 M ammonium acetate (80/20) and methanol/ethyl acetate (70/30) as eluents. Chromatographic peaks were detected using a UV-VIS photodiode array detector (Beckman 168) and identified based on retention time and comparison with absorbance spectra of known pigment standards (Sigma Chemical Company, MO, USA; and the DHI Institute for Water and Environment, Hørsholm, Denmark). Divinyl chlorophyll-a and divinyl chlorophyll-b, characteristic of *Prochlorococcus* sp., were quantified by acidifying samples with 1 N HCL and recording the peak heights of divinyl phaeophytin-like pigments. The quality of pigment data was strictly controlled following the method described by Aiken et al. [33]. After that, 1392 samples combined with the phytoplankton absorption data were retained for this study.

In the following, the term "pigment" is used for either a specific type of pigment or a pigment group such as photosynthetic carotenoids (PSC) and PPC. The essential information of pigments used in this study is listed in Table 1.

Table 1. Name, abbreviation, and statistical distribution of pigments used in this study. Statistical distribution includes maximum (Max., mg m^{-3}), minimum (Min., mg m^{-3}), and standard deviation (std., mg m^{-3}) of pigment concentrations. Pearson correlation coefficient (*r*) is estimated for linear regression between log-transformed concentrations of individual pigments and total chlorophyll-a. All statistics were calculated for values >0 mg m^{-3}. Values of pigment concentrations below the limit of detection were considered zero.

Name	Abbreviation	Min.–Max. (Std.)	*r*	N
chlorophyll-a	Chl-a [1]	0.03–28.0 (2.41)	1.00	1392
alloxanthin	Allo	0.001–0.24 (0.03)	0.34	765
19′-butanoyloxy-fucoxanthin	But19	0.001–0.57 (0.05)	0.49	762
chlorophyll b	Chl-b [2]	0.001–0.62 (0.07)	0.47	1185
chlorophyll-c_1+ chlorophyll-c_2	Chl-$c_{1,2}$ [3]	0.001–5.37 (0.48)	0.95	1364
chlorophyll-c_3	Chl-c_3	0.001–0.39 (0.04)	0.56	1036
chlorophyll c	Chl-c [4]	0.001–5.52 (0.50)	0.94	1366
diadinoxanthin	Diadino	0.001–1.06 (0.11)	0.88	1301
fucoxanthin	Fuco	0.001–16.6 (1.38)	0.91	1325
19′-hexanoyloxy-fucoxanthin	Hex19	0.002–1.39 (0.13)	0.43	1080
peridinin	Perid	0.003–0.53 (0.08)	0.43	489
zeaxanthin	Zea	0.001–0.21 (0.03)	0.04	626
photosynthetic carotenoids	PSC [5]	0.008–16.6 (1.35)	0.97	1392
photoprotective carotenoids	PPC [6]	0.005–1.45 (0.13)	0.84	1368

[1] Chl-a = chlorophyllide-a + mono- and divinyl chlorophyll-a + divinyl chlorophyll-a + chlorophyll-a epimer and allomer (mg m^{-3}). [2] Chl-b = chlorophyll-b + dv-Chl-b (mg m^{-3}). [3] Chl-$c_{1,2}$ = chlorophyll-c_1 + chlorophyll-c_2 (mg m^{-3}). [4] Chl-c = Chl-c_1 + Chl-c_2 + Chl-c_3 (mg m^{-3}). [5] PSC = α-carotene + 19′-hexanoyloxyfucoxanthin + 19′-butanoyloxyfucoanthin + fucoxanthin + peridinin + prasinoxanthin (mg m^{-3}). [6] PPC = β-Carotene + zeaxanthin + diadinoxanthin + diatoxanthin + alloxanthin (mg m^{-3}).

2.2. Model Development

As illustrated in Figure 2, three model types were examined in this study to estimate pigment concentrations based on Chl-a concentration or phytoplankton absorption spectra. Various parameters were applied to compare the performance of each of the three model

types. The same 75% of all samples were used as a training set for building the models and the same 25% of all samples were used as a testing set for validating the models.

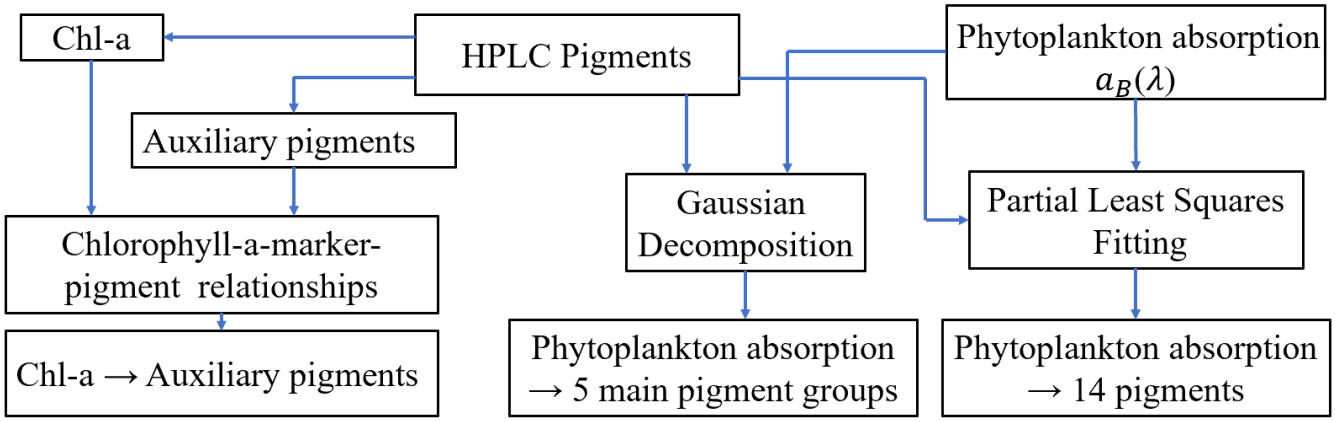


Figure 2. Schematic overview of the data and processing steps used to estimate pigments from phytoplankton absorption and total chlorophyll-a (Chl-a) concentration.

2.2.1. Chl-a Model

The Chl-a model formula can be expressed as follows, which is an adjusted form of that used in Chase et al. [7]:

$$C_j = A_j \times {C_1}^{B_j}, \quad (1)$$

where C_j is the value of the j-th predicted pigment concentration, C_1 is the observed concentration of Chl-a, and A_j and B_j are model coefficients associated with the j-th pigment, determined using the training set. The MATLAB function "fit" was used for calculating the coefficients of the model, and "NonlinearLeastSquares" was the selected option for fitting.

2.2.2. Gaussian Model

The principle of Gaussian decomposition is that a phytoplankton absorption spectrum can be segmented typically into 11–13 Gaussian bands, each of which would be associated with a specific pigment group, and that the concentration of that pigment group could be estimated from the amplitude of the Gaussian band.

We used the method of Hoepffner and Sathyendranath [21], as implemented by Chase et al. [5] and Ye et al. [25], in which an optimized fixed-parameter configuration was selected for the central wavelength of each Gaussian band and its width. Each spectrum was decomposed by optimizing the overdetermined equation, which expressed the absorption spectrum as the sum of the individual absorption bands, by minimizing the error between the observed and reconstructed absorption spectra:

$$\chi^2 = \sum_{\lambda=400}^{700} \left(a_{\mathrm{B}}(\lambda) - \left(\sum_{i=1}^{13} a_{\mathrm{g}}(\lambda_{\mathrm{m}i}) \exp\left\{ -0.5 \left[\frac{\lambda - \lambda_{\mathrm{m}i}}{\sigma_i} \right]^2 \right\} \right) \right)^2, \quad (2)$$

where $a_{\mathrm{B}}(\lambda)$ is the measured phytoplankton absorption, $a_{\mathrm{g}}(\lambda_{\mathrm{m}i})$ denotes the amplitude of the i-th Gaussian function, and $\lambda_{\mathrm{m}i}$ and σ_i represent the central wavelength with maximum absorption and the width of the i-th Gaussian band, respectively. The range of wavelengths (400–700 nm) was treated at 1-nm increments. Here, the MATLAB function "lsqnonneg" was used to solve the nonlinear least square problem in Equation (2) and to invert the amplitudes a_g.

The fixed central wavelengths (peak wavelengths) of the 12 Gaussian bands and their widths, together with the pigments associated with each band, are presented in

Table 2. Although biliproteins were not analyzed in this work, one peak was assigned to phycoerythrin to ensure good fits.

Table 2. Central wavelengths of the Gaussian peak and the widths (σ) of the Gaussian functions are assigned for different phytoplankton pigments (PE = phycoerythrin).

	Chl-a	**Chl-a**	**Chl-c**	**Chl-b**	**PPC**	**PSC**	**PE**	**Chl-c**	**Chl-a**	**Chl-c**	**Chl-b**	**Chl-a**
Peak (nm)	406	434	453	470	492	523	550	584	617	638	660	675
σ (nm)	17	13	12	14	17	15	15	17	14	11	11	10

A power function was used to model the relationship between the magnitude of the Gaussian peak and the concentration of the associated pigment, as reported in Chase et al. [7]:

$$C_j = A_{2,i,j} \times a_g(\lambda_{mi})^{B_{2,i,j}}, \quad (3)$$

where C_j is the value of the j-th predicted pigment concentration, and $a_g(\lambda_{mi})$ is the value of the i-th Gaussian peak at its central wavelength, λ_{mi}. The coefficients $A_{2,i,j}$ and $B_{2,i,j}$ are the fitted parameters. The MATLAB function is used to calculate A_2 and B_2 was the same as that mentioned in Section 2.2.1.

2.2.3. Partial Least Squares Model

In the PLS model, principal component analysis and canonical correlation analysis were applied to phytoplankton absorption data (predictor variables, Xv) and pigments (response variables, Yv). First, the analyses were repeated for Xv and Yv to decide the required number of principal components (PCN) that maximize the covariance of Xv and Yv and maximize the correlation between Xv and Yv simultaneously. Details of the steps adopted for the selection of PCN are given in the following.

- Step 1: For each analysis, all absorption samples with full wavelengths were split into two parts randomly, 75% of the dataset was used for training, and the remaining 25% was used for validation.
- Step 2: Typically, the first 15 or fewer principal components (PCs) explained more than 99% of the total variance of the absorption spectrum, which demonstrates that the first 15 or fewer PCs contain most of the signals of the absorption spectrum. The explained variance for the independent and dependent variables was a valuable parameter derived from function "plsregress" in MATLAB. Therefore, the analyses were limited to the first 15 PCs. Within the first 15 PCs, the PCs were added sequentially from number 2. The weights associated with the corresponding PCs in the training dataset were regressed against the corresponding pigment concentration based on a leave-one-out cross-validation (CV) calculation. The coefficient of determination (r^2) of the PLS regression between each pigment and the corresponding principal component was tabulated.
- Step 3: Steps 1 and 2 were repeated 100 times, and the mean r^2 and the mean value of the root mean square error (RMSE) were calculated for each pigment–PCN pair (Figure 3).
- Step 4: High RMSEs implied that the PCN was unsuitable for building the relationship between the PCs and pigment concentration. The ideal pigment–PCN pair would therefore have a high r^2 value and a low RMSE. We thus identified the optimal number of PCs necessary for the analysis based on the maximum r^2 and low standard RMSE.

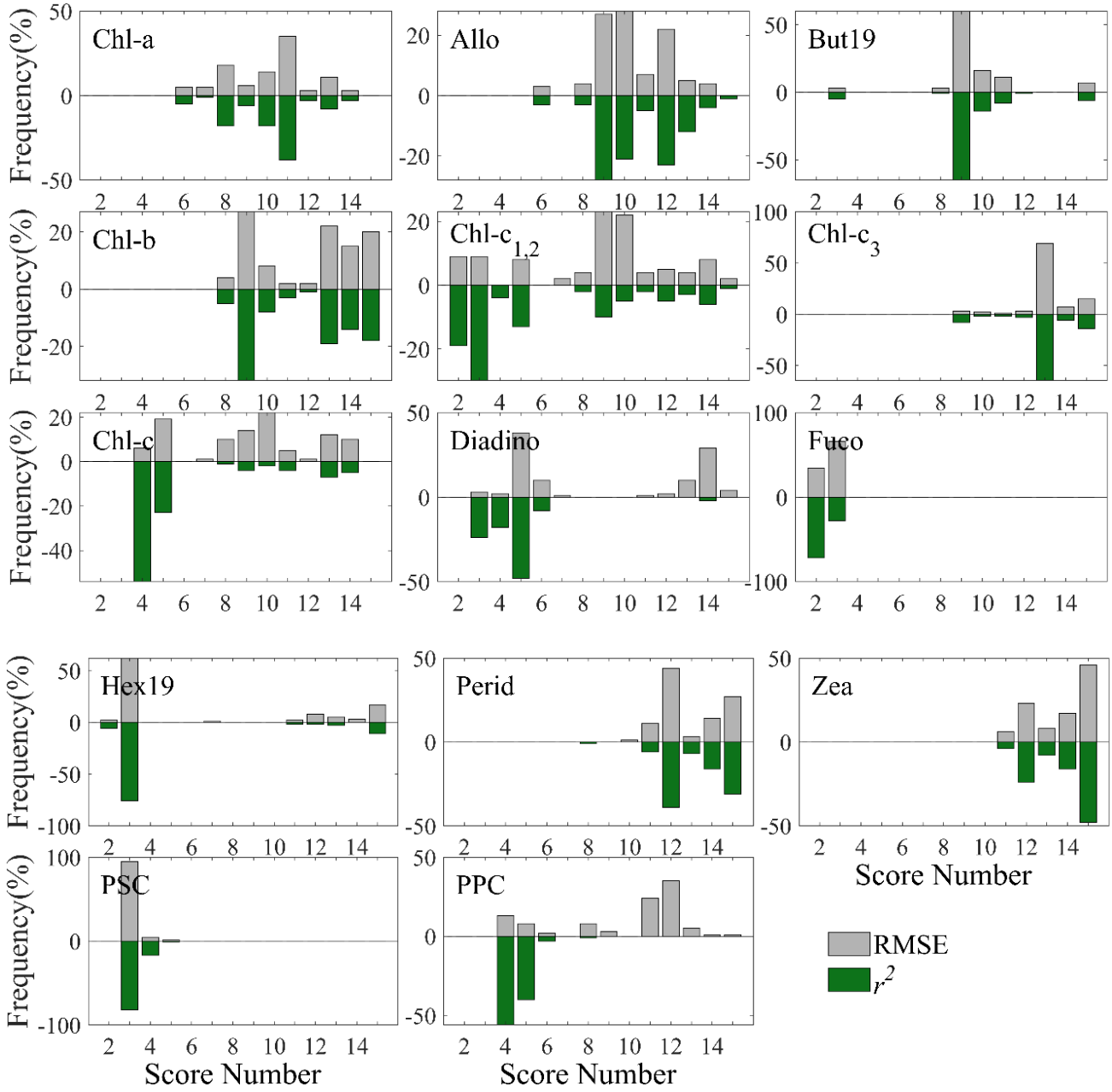


Figure 3. Example of selection of a number of principal components calculated based on $a_B(\lambda)$. The frequency of the minimum RMSE (gray column) and the frequency of the maximum r^2 (green column) was calculated from 100 permutations for each pigment. The negative sign simply represents another type of parameter.

Following the determination of the PCN, the inversion model could be established using the built-in function "plsregress" provided in MATLAB. The inputs for the PLS model were absorption spectra, target pigment concentration, and PCN related to the specific pigment. The main outputs of the PLS model were the corresponding coefficients for retrieving pigment concentrations.

Derivative spectra (i.e., second and fourth derivatives of phytoplankton absorption spectra, $a''_B(\lambda)$ and $a^{4th}_B(\lambda)$, respectively) of the phytoplankton absorption spectra were used to identify the major absorption signatures and exploit the utility for retrieving pigment concentrations. The derivative of $a_B(\lambda)$ was estimated using a finite difference approximation, as recommended by Organelli et al. [23], with a band separation of 9 nm after smoothing [8].

Selected multispectral bands of $a_B(\lambda)$ were also used as inputs to the PLS model. The bands were selected based on typical ocean-colour satellite sensors (i.e., SeaWiFS, MODIS, and MERIS).

2.3. Model Assessment

The Pearson correlation coefficients (r), r^2, RMSE, bias (Bias), and standard deviation (std) values were all calculated using the log-transformed predicted pigment concentration (C_{pi}) and log-transformed observed pigment concentration (C_{oi}). The median absolute percent difference (MAPD) was calculated using the non-log-transformed pigment concentration. We discussed the results obtained using the different models from the following three perspectives:

1. Training model. Calculating statistical parameters based on the leave-one-out CV method [8,23]. The model was trained using the $k-1$ sample selected from the training set (k samples) and verified using the k-th sample. After all, samples were traversed, k model results could be obtained from the model evaluation.
2. Validating model. Statistical parameters were calculated by putting the testing set into the model built with the training set.
3. Validating model with CV calculation. Statistical parameters were calculated based on the CV for 500 permutations [6,34,35]. The training set was divided into two parts randomly, as mentioned in Section 2.2.3. For each pigment, the model was fitted using S_train and validated using S_test for the 500 permutations, following which we obtained 500 groups of evaluation parameters. For each parameter, the mean of the 500 results was used as the final indicator of the model.

Parameters without "CV" were calculated as follows:

$$\text{RMSE} = \text{Root of } \frac{1}{N}\sum_{i=1}^{N}\left(\ln(C_{oi}) - \ln\left(C_{pi}\right)\right)^2, \tag{4}$$

$$\text{Bias} = \frac{1}{N}\sum_{i=1}^{N}\left(\ln(C_{pi}) - \ln(C_{oi})\right), \tag{5}$$

$$\text{MAPD} = \text{Median}\left(\frac{\left|C_{oi} - C_{pi}\right|}{C_{oi}} \times 100\right). \tag{6}$$

Parameters with "CV" were calculated as follows:

$$\text{RMSE}_{CV} = \text{Root of } \frac{1}{N}\sum_{i=1}^{N}\left(\ln\left(C_{oi}^{val}\right) - \ln\left(C_{pi}^{val}\right)\right)^2, \tag{7}$$

$$\text{Bias}_{CV} = \frac{1}{N}\sum_{i=1}^{N}\left(\ln\left(C_{oi}^{val}\right) - \ln\left(C_{pi}^{val}\right)\right), \tag{8}$$

$$\text{MAPD}_{CV} = \text{Median}\left(\frac{\left|C_{oi}^{val} - C_{pi}^{val}\right|}{C_{oi}^{val}} * 100\right), \tag{9}$$

where C_o^{val} and C_p^{val} are the observed and predicted concentrations selected or retrieved from S_test, respectively, and N is the number of samples used in the model.

3. Results

3.1. Chl-a Model

As shown in Table 3 and Figure 4, the retrieval performance for Chl-$c_{1,2}$, chlorophyll c (Chl-c), diadinoxanthin (Diadino), fucoxanthin (Fuco), PSC, and PPC derived using the Chl-a model was better than that for other pigments, with r^2 values of >0.7, RMSEs of <0.53 mg m^{-3}, and Bias of <0.18 mg m^{-3}, except for Fuco. Predicted and measured data points of these pigments were distributed evenly around the 1:1 line. A high r^2 value corresponds to a slope (S) close to 1 and an intercept (I) close to 0, which can be observed in Figure 4. Relatively low values of their MAPD also suggest that the Chl-a model could

retrieve these pigments reliably. With consideration of the analysis in Table 1, it can be ascertained that the correlation coefficients between these pigments and Chl-a were all >0.8, suggesting that these pigments have reasonably strong empirical relationships with Chl-a.

Table 3. Statistics for the Chl-a model including A and B in Equation (1), slope (*S*), intercept (*I*), and determination coefficient (r^2) for the linear regression, root mean square error (RMSE, mg m^{-3}), median absolute percent difference (MAPD, %), and bias (Bias, mg m^{-3}).

Pigment	For Training Set					For Testing Set			
	A	**B**	r^2	*S*	*I*	r^2	**RMSE**	**MAPD**	**Bias**
Allo	0.04	0.33	0.14	0.07	−3.14	0.04	0.88	54.0	0.34
But19	0.06	0.61	0.22	0.40	−1.80	0.29	0.70	29.1	0.14
Chl-b	0.09	0.26	0.23	0.16	−1.97	0.20	0.76	40.82	0.27
Chl-$c_{1,2}$	0.16	1.07	0.91	0.82	−0.22	0.91	0.47	24.9	0.17
Chl-c_3	0.03	0.49	0.32	0.30	−2.30	0.29	0.86	44.6	0.34
Chl-c	0.18	1.04	0.89	0.80	−0.26	0.89	0.48	27.7	0.14
Diadino	0.07	0.77	0.76	0.69	−0.71	0.79	0.52	39.3	0.17
Fuco	0.41	1.09	0.82	0.69	0.04	0.85	0.81	62.2	0.48
Hex19	0.17	0.33	0.25	0.10	−1.68	0.07	0.98	50.0	0.36
Perid	0.09	0.49	0.27	0.07	−2.23	0.02	0.96	46.5	0.36
Zea	0.04	0.07	0.00	0.00	−3.37	0.01	0.91	60.8	0.39
PSC	0.57	0.96	0.93	0.86	−0.01	0.94	0.34	17.6	0.11
PPC	0.12	0.59	0.68	0.64	−0.70	0.74	0.47	31.6	0.12

Values of RMSE, Bias, and MAPD for Fuco were higher than for the other pigments, which might reflect the poor performance of the model for lower concentration samples of Fuco (Figure 4). Moreover, other pigments such as alloxanthin (Allo), Chl-b, chlorophyll-c_3 (Chl-c_3), 19′-hexanoyloxy-xanthin (Hex19), peridinin (Perid), and zeaxanthin (Zea) are hardly retrieved by the Chl-a model. Specifically, all these pigments were estimated with values of $r^2 < 0.29$, Bias > 0.27 mg m^{-3}, and RMSE > 0.76 mg m^{-3}. The slopes of the data trend for these pigments were all <0.2, as reflected by the scattered points in Figure 4. Although there was an obvious trend along the 1:1 line for 19′-butanoyloxy-fucoxanthin (But19), the retrieved points were more dispersed than those for other pigments such as Chl-c. Coincidentally, the correlations between the concentrations of these pigments and Chl-a were all <0.56 (Table 1), suggesting weak relationships between these pigments and Chl-a.

3.2. Gaussian Model

As shown in Table 4, large differences existed in the model coefficients for different pigments. Specifically, the A_2 coefficient varied widely among the models based on different wavebands for Chl-a inversion. Validation results calculated from the testing set for Chl-a, Chl-b, Chl-c, PSC, and PPC showed consistency with those calculated from the training set (Table 4). For Chl-a, the highest value of r^2 and the lowest values of RMSE and MAPD were all obtained at 675 nm. The r^2 value of Chl-b, calculated either from the 470-nm band or from the 660-nm band, was lower than that of other pigments. The sample points of Chl-b deviated obviously from the 1:1 line (Figure 5) and the slope of the Chl-b inversion was far from the value of 1 (Table 4). The inversion quality obtained for Chl-c was similar at 584 and 638 nm with $r^2 > 0.7$. A clear 1:1 trend could be found for Chl-c despite the higher values of RMSE, MAPD, and Bias in Table 4. Satisfactory performance was also obtained for PSC and PPC at 523 and 492 nm, respectively. The r^2 value of PPC was slightly lower than that of PSC, but the values of RMSE, MAPD, and Bias for PPC were lower than those for PSC. The above results indicate that the Gaussian model performed well for Chl-a, Chl-c, PSC, and PPC but not for Chl-b.

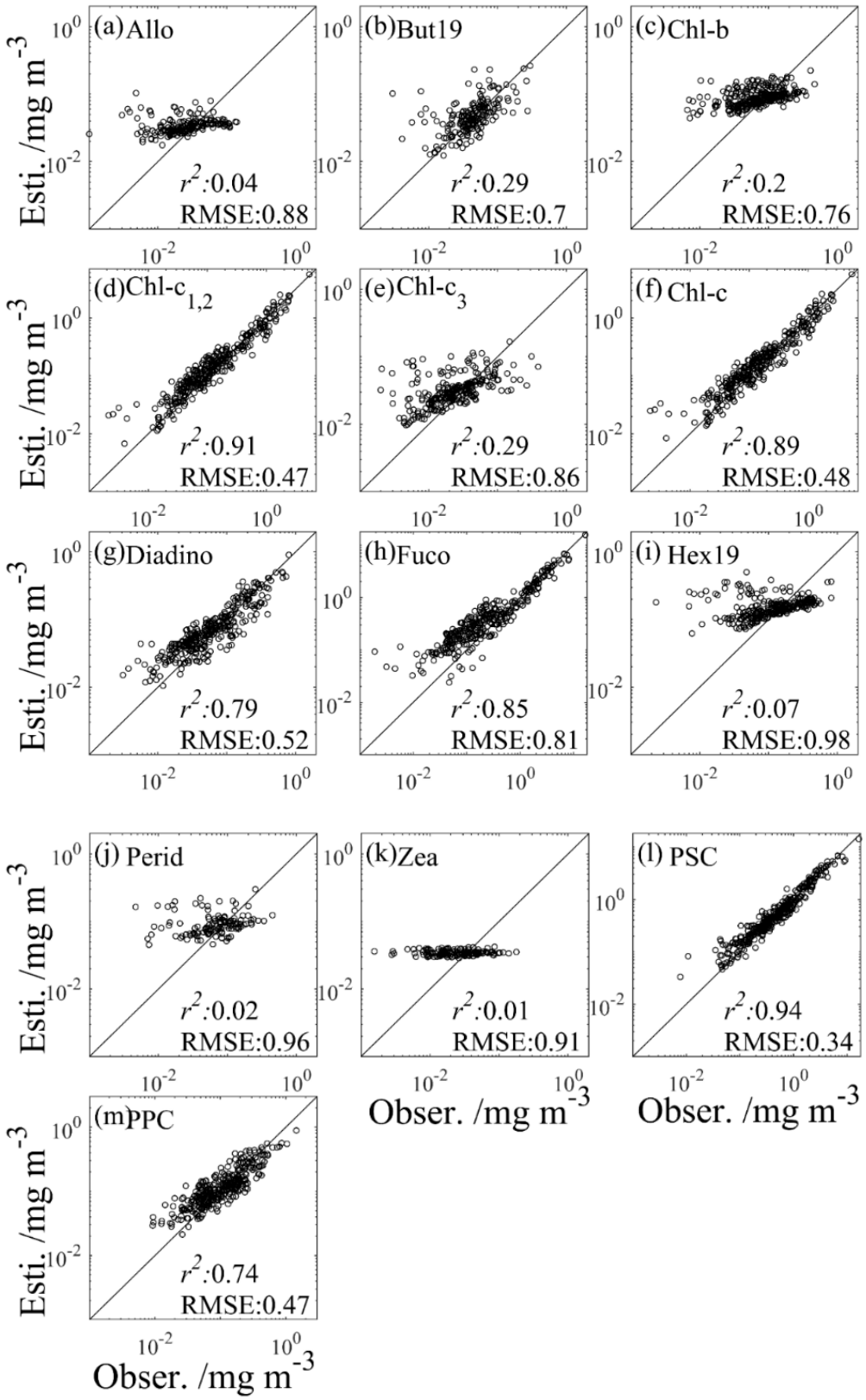


Figure 4. Observed concentrations (Obser.) versus concentrations estimated from the Chl-a model (Esti.) of 13 phytoplankton pigments (**a**–**m**) using the testing set. The black line represents the 1:1 line.

Table 4. Statistics for the Gaussian model including A_2 and B_2 for Equation (3), slope (*S*), intercept (*I*), and determination coefficient (r^2) for the linear regression, root mean square error (RMSE, mg m^{-3}), median absolute percent difference (MAPD, %), and bias (Bias, mg m^{-3}).

Pigment	For Training Set					For Testing Set			
	A_2	B_2	r^2	*S*	*I*	r^2	RMSE	MAPD	Bias
Chl-a_406	32.8	1.05	0.70	0.77	0.18	0.82	0.53	39.5	0.22
Chl-a_434	59.5	1.21	0.71	0.79	0.11	0.76	0.57	41.7	0.15
Chl-a_617	229	1.01	0.73	0.87	0.11	0.85	0.45	28.1	0.13
Chl-a_675	67.0	1.08	0.84	0.89	0.10	0.86	0.44	29.4	0.12
Chl-b_470	0.6	0.51	0.30	0.29	−1.66	0.35	0.68	39.9	0.23
Chl-b_660	0.4	0.28	0.28	0.17	−1.96	0.20	0.75	43.5	0.25
Chl-c_584	56.9	1.10	0.67	0.71	−0.29	0.75	0.73	45.4	0.28
Chl-c_638	56.3	1.13	0.74	0.73	−0.29	0.74	0.74	44.1	0.25
PSC_523	80.4	1.10	0.66	0.79	0.05	0.84	0.55	34.1	0.23
PPC_492	3.1	0.89	0.66	0.67	−0.66	0.70	0.49	31.2	0.10

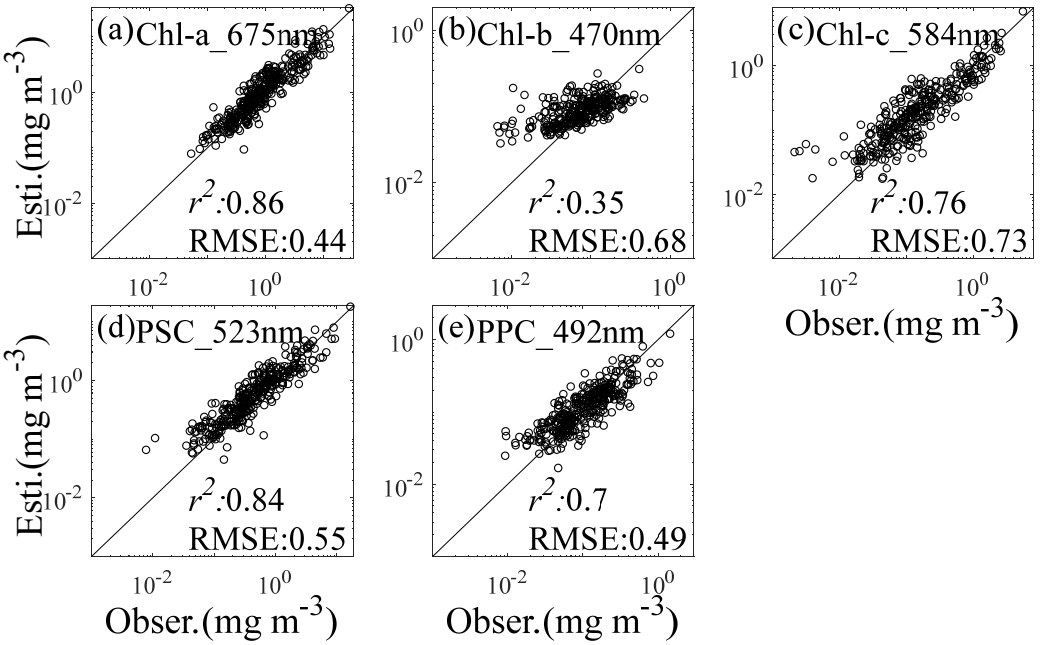


Figure 5. Observed concentrations (Obser.) versus concentrations estimated from the Gaussian model (Esti.) of five main phytoplankton pigment groups (**a**–**e**) using the testing set. The black line represents the 1:1 line.

3.3. PLS Model

3.3.1. Analysis of Inversion Results

In this section, we analyze only the PLS models based on original data and full hyperspectral data; the models corresponding to derivative data or multispectral data are discussed in Sections 3.3.2 and 3.3.3. Results showed that the explained variances for the independent values were >99.8% for all pigments, suggesting excellent capability for signal extraction using the selected PCN. The explained variances for the dependent parameters showed lower values and large differences in performance among these pigments. It can be seen in Table 5 and from Figure 6 that the points of some pigments were scattered in the low-concentration region, but compact in the high-value region, e.g., Chl-a, Chl-c_3, and Fuco. Specifically, the values of MAPD for all pigments were relatively low (24%–49%), while the RMSE values for But19, Chl-c_3, Fuco, Perid, and Zea were slightly higher (RMSE > 0.6 mg m^{-3}). Values of r^2 for Chl-a, Chl-b, Chl-$c_{1,2}$, Chl-c, Diadino, Fuco, PSC, and PPC were higher than those for other pigments. Overall, the inversion performances

of Chl-a, Chl-b, Chl-$c_{1,2}$, Chl-c, PSC, and PPC were better than those of other pigments considering the r^2, RMSE, and Bias values. The other pigments had relatively low values of r^2 but clear distribution of points (Figure 6) around the 1:1 line. Reasonable consistency could be found among the slopes of the data trends, r^2 obtained from the training set, and those calculated from the testing set, suggesting that PLS is a promising and stable method for retrieval of concentrations of various pigments.

Table 5. Statistics for the PLS model including the explained variance for independent (X, %) and dependent (Y, %) variables corresponding to the relative PCN (see Table 6), slope (*S*), intercept (*I*), determination coefficient (r^2) for the linear regression, root mean square error (RMSE, mg m^{-3}), median absolute percent difference (MAPD, %), and bias (Bias, mg m^{-3}).

Pigment	For Training Set					For Testing Set			
	X	**Y**	r^2	***S***	***I***	r^2	**RMSE**	**MAPD**	**Bias**
Chl-a	99.99	91.51	0.77	0.82	0.05	0.81	0.50	27.0	0.08
Allo	99.99	69.40	0.48	0.59	−1.41	0.44	0.66	26.0	0.12
But19	99.99	48.54	0.22	0.38	−1.86	0.28	0.70	25.7	0.13
Chl-b	99.99	69.68	0.71	0.69	−0.76	0.67	0.45	23.6	0.08
Chl-$c_{1,2}$	99.83	89.87	0.76	0.82	−0.29	0.77	0.68	34.1	0.10
Chl-c_3	99.99	72.67	0.56	0.63	−1.37	0.45	0.74	43.1	0.05
Chl-c	99.91	90.19	0.75	0.73	−0.38	0.78	0.65	33.8	0.15
Diadino	99.91	73.65	0.52	0.69	−0.81	0.62	0.66	34.1	0.08
Fuco	99.84	81.17	0.75	0.82	−0.02	0.72	0.86	48.7	0.23
Hex19	99.84	52.84	0.50	0.47	−1.03	0.50	0.69	33.5	0.17
Perid	99.99	62.41	0.52	0.59	−0.97	0.49	0.63	27.8	0.14
Zea	99.99	58.85	0.38	0.57	−1.47	0.51	0.60	38.0	0.15
PSC	99.84	80.92	0.82	0.83	−0.04	0.83	0.51	28.6	0.11
PPC	99.88	68.73	0.77	0.65	−0.70	0.73	0.48	32.5	0.11

Table 6. The PCN for each pigment when different types of spectra are used, which includes $a_B(\lambda)$, $a''_B(\lambda)$, $a_B^{4th}(\lambda)$, Multi_SeaWiFS, Multi_MODIS, and Multi_MERIS.

Pigment	Hyperspectral			Multispectral		
	$a_B(\lambda)$	$a''_B(\lambda)$	$a_B^{4th}(\lambda)$	**Multi_SeaWiFS**	**Multi_MODIS**	**Multi_MERIS**
Chl-a	11	4	3	3	2	5
Allo	12	10	11	5	6	7
But19	9	8	8	5	5	5
Chl-b	9	8	15	6	7	7
Chl-$c_{1,2}$	3	3	4	5	5	3
Chl-c_3	13	9	15	5	6	5
Chl-c	4	3	3	5	5	5
Diadino	5	2	3	4	7	4
Fuco	3	2	3	4	4	4
Hex19	3	3	3	3	3	6
Perid	12	11	15	3	7	8
Zea	15	15	7	4	7	8
PSC	3	2	3	2	4	4
PPC	4	4	3	6	3	6

3.3.2. Comparison of Inversion Methods Using Absorption Spectra and Their Derivative Spectra

Derivative spectral analysis is usually used to analyze high-resolution and spectrally continuous remote sensing data [22,36]. To determine the advantage of using the derivative of phytoplankton absorption spectra for retrieving pigment concentrations, we compared the performances of the PLS models trained with $a_B(\lambda)$, $a''_B(\lambda)$, and $a_B^{4th}(\lambda)$. The first derivative of the spectra was not selected because it could weaken the signal of absorption

to a value of zero. The corresponding optimal PCN for each type of spectra is listed in Table 6, and the CV results of the model performances are shown in Figure 7.

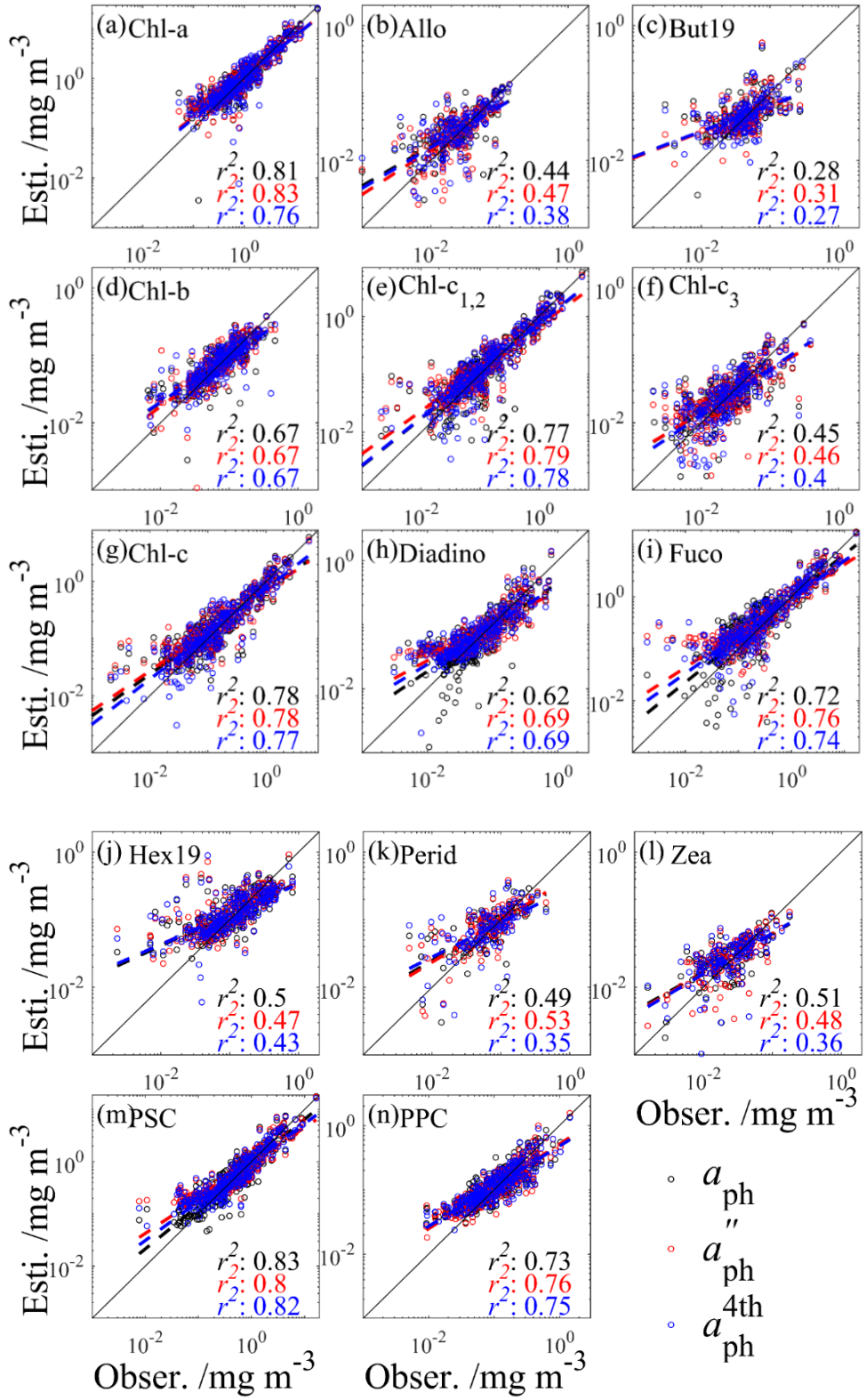


Figure 6. Comparison of inversion performances derived by the PLS model of 14 phytoplankton pigments (**a**–**n**) using $a_B(\lambda)$ (black), $a''_B(\lambda)$ (red), and $a^{4th}_B(\lambda)$ (blue) as input. The best-fit line shows a first-order linear regression performed on the log-transformed data. The black line represents the 1:1 line. Results were obtained using the testing set together with the model tuned using the training set.

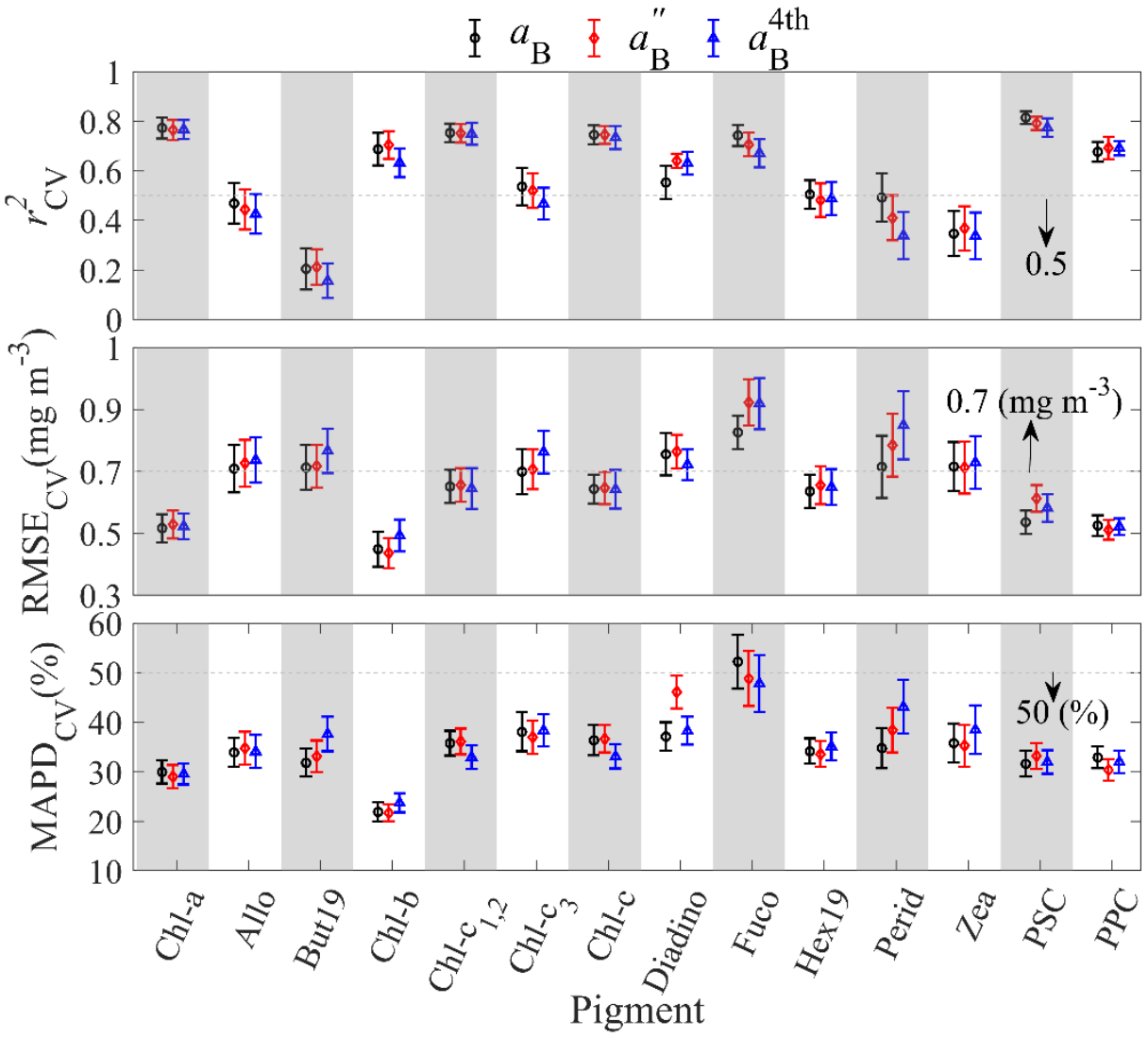


Figure 7. Cross-validation (CV) statistics (r^2_{CV}, $RMSE_{CV}$, $MAPD_{CV}$) between observed pigment concentrations and modeled pigment concentrations produced by the PLS model. Results for $a_B(\lambda)$ (black), $a''_B(\lambda)$ (red), and $a^{4th}_B(\lambda)$ (blue) are compared.

As shown in Figure 7, a lower r^2_{CV} value was generally obtained from $a^{4th}_B(\lambda)$ for most pigments except Diadino. Values of $RMSE_{CV}$ derived from $a^{4th}_B(\lambda)$ were slightly higher than those derived from $a_B(\lambda)$ and $a''_B(\lambda)$ for But19, Chl-c$_3$, Fuco, and Perid. For these pigments, values of $MAPD_{CV}$ were also higher for But19 and Perid computed based on $a^{4th}_B(\lambda)$. By contrast, the performance of both $a_B(\lambda)$ and $a''_B(\lambda)$ was better than that of $a^{4th}_B(\lambda)$ for most pigments, and there was generally little difference between the performances of $a_B(\lambda)$ and $a''_B(\lambda)$. There was also little difference between the standard deviations of r^2_{CV}, $RMSE_{CV}$, and $MAPD_{CV}$ for all pigments among the different derivative spectra.

3.3.3. Comparison of Inversion Performance of Hyperspectral Absorption and Multispectral Absorption

To investigate whether hyperspectral data could improve the retrieval accuracy in comparison with limited spectral bands, we compared the inversion performances of the PLS models by inputting hyperspectral or multiband $a_B(\lambda)$ to the models as independent variables. Multispectral $a_B(\lambda)$ corresponding to SeaWiFS bands, MODIS bands, and MERIS bands were denoted as Multi_SeaWiFS, Multi_MODIS, and Multi_MERIS, respectively.

In our study, the retrieval accuracy of Chl-a, But19, Chl-c$_{1,2}$, Chl-c, Diadino, Fuco, Hex19, PSC, and PPC obtained using hyperspectral data and multispectral data was similar

according to the values of r^2_{CV}, $RMSE_{CV}$, and $MAPD_{CV}$; the $MAPD_{CV}$ of Fuco calculated using multispectral data was particularly low. For other pigments, including Allo, Chl-b, Chl-c_3, Perid, and Zea, the hyperspectral absorption spectra had an obvious advantage that was reflected in the values of r^2_{CV}, $RMSE_{CV}$, and $MAPD_{CV}$ (Figure 8).

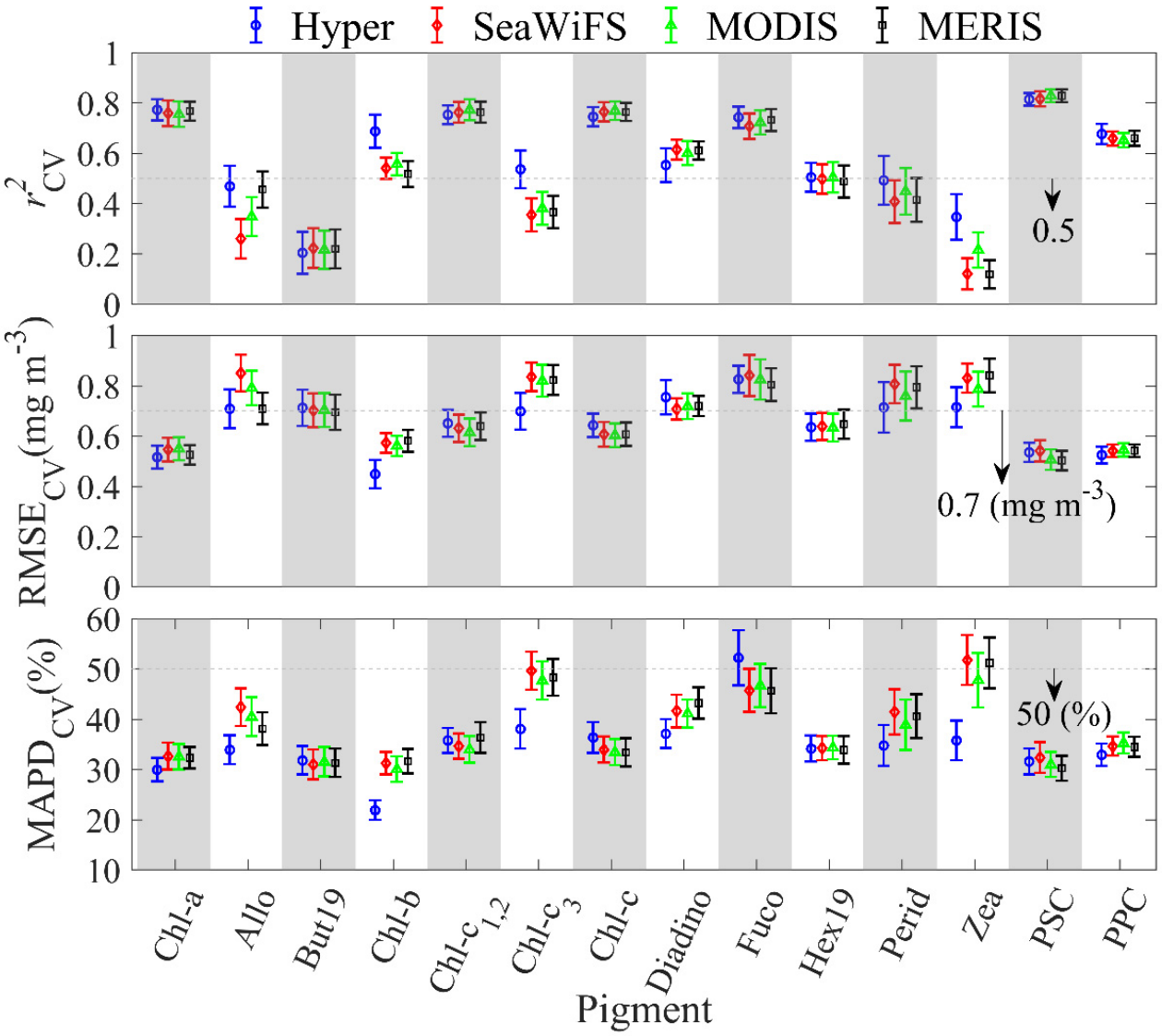


Figure 8. Cross validation (CV) statistics (r^2_{CV}, $RMSE_{CV}$, and $MAPD_{CV}$) between observed pigment concentrations and modeled pigment concentrations produced by the PLS model using hyperspectral $a_B(\lambda)$, Multi_SeaWiFS, Multi_MODIS, and Multi_MERIS.

The performances achieved using multispectral data showed certain differences. Better performance for Allo was obtained using the MERIS bands in comparison with the performance achieved with the bands of the other two sensors (Figure 8). The MERIS bands differed from those of SeaWiFS and MODIS at 413, 560, 620, 665, and 681 nm, suggesting that the SeaWiFS and MODIS bands do not carry the same spectral information for Allo as carried by the MERIS bands.

3.4. Comparison of the Three Model Types

The performances of the three model types can be compared by considering two aspects: retrieved pigment types and model accuracy. Taylor diagrams (Figure 9) were

used to compare the relative precision of the different models, and the CV results (Table 7) are presented here.

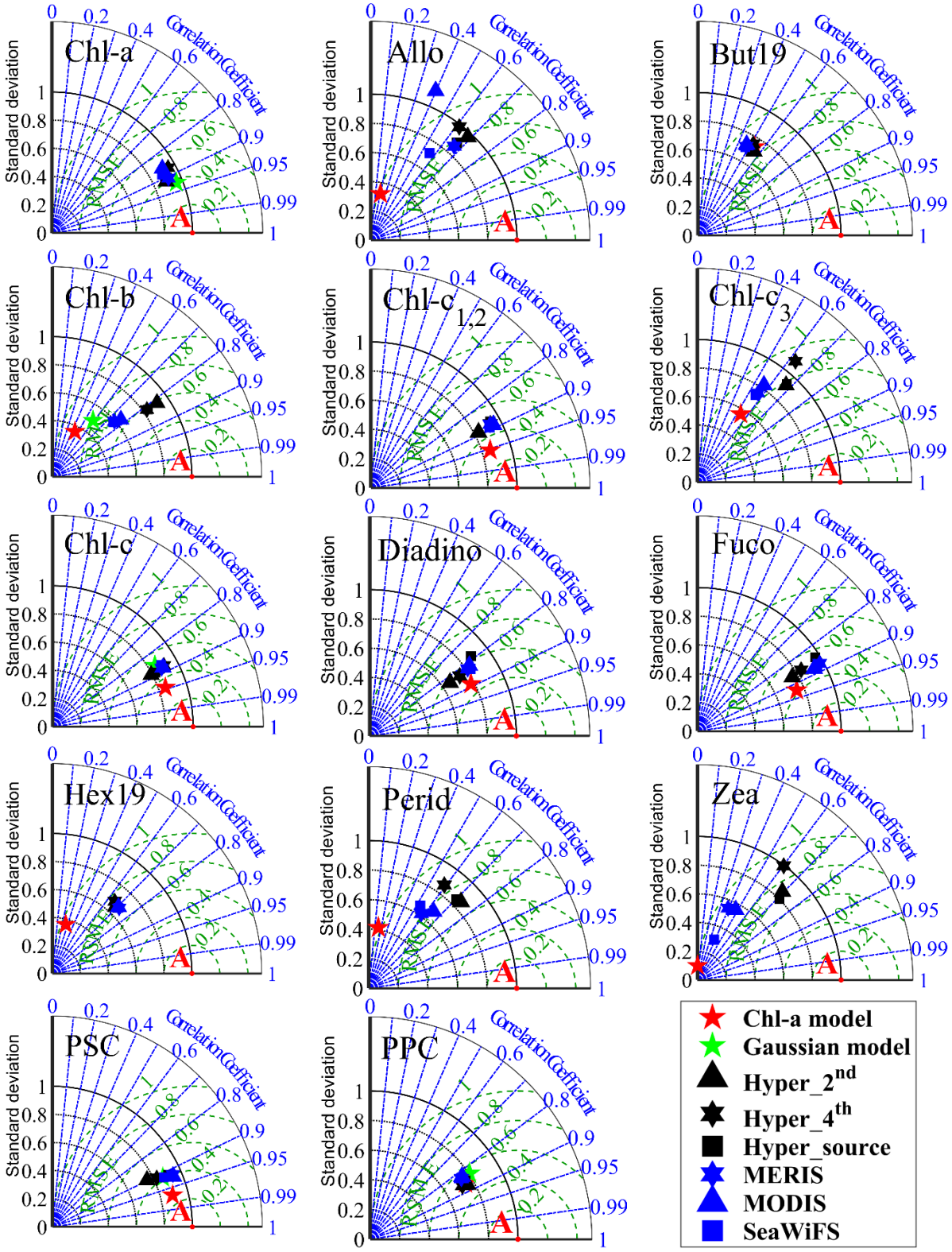


Figure 9. Taylor diagrams illustrating the relative model performances of the Chl-a model, Gaussian model, and PLS model (based on $a_B(\lambda)$, $a''_B(\lambda)$, $a_B^{4th}(\lambda)$, Multi_SeaWiFS, Multi_MODIS, and Multi_MERIS). The correlation coefficient, normalized standard deviation, and normalized RMSE are presented in the Taylor diagrams. The red dot labeled "A" represents perfect model performance. A model performs better if the retrievals are closer to the red dot A. All parameters were calculated based on log-transformed pigment concentration.

Table 7. Mean statistics of the CV of 500 permutations based on S_train and S_test for the three model types, i.e., the Chl-a model, Gaussian model, and PLS model based on hyper $a_B(\lambda)$. For the Gaussian model, the optimum performance for each pigment was chosen for presentation here.

Pigment	r^2_{CV}	$RMSE_{CV}$	$MAPD_{CV}$	$Bias_{CV}$	r_{CV}
		Chl-a model			
Allo	0.16	0.94	55.20	0.37	0.39
But19	0.22	0.70	31.70	0.13	0.47
Chl-b	0.23	0.75	39.97	0.24	0.47
Chl-$c_{1,2}$	0.91	0.42	24.64	0.15	0.95
Chl-c_3	0.33	0.89	50.67	0.35	0.57
Chl-c	0.89	0.42	27.37	0.11	0.95
Diadino	0.76	0.61	39.26	0.26	0.87
Fuco	0.82	0.84	57.34	0.49	0.91
Hex19	0.25	0.83	47.67	0.32	0.50
Perid	0.27	0.95	46.35	0.36	0.52
Zea	0.01	0.89	53.44	0.30	0.06
PSC	0.93	0.34	19.71	0.13	0.97
PPC	0.69	0.53	32.18	0.17	0.83
		Gaussian model based on $a_B(\lambda)$			
Chl-a_675	0.83	0.47	30.64	0.15	0.91
Chl-b_470	0.41	0.66	35.45	0.21	0.64
Chl-c_584	0.71	0.71	42.97	0.25	0.84
PSC_523	0.81	0.57	37.18	0.25	0.90
PPC_492	0.66	0.53	34.25	0.14	0.81
		PLS model based on $a_B(\lambda)$			
Chl-a	0.77	0.52	29.95	0.11	0.88
Allo	0.47	0.71	33.91	0.12	0.68
But19	0.20	0.71	31.81	0.14	0.44
Chl-b	0.69	0.45	21.91	0.08	0.83
Chl-$c_{1,2}$	0.75	0.65	35.78	0.16	0.87
Chl-c_3	0.54	0.70	38.05	0.14	0.73
Chl-c	0.75	0.64	36.37	0.19	0.86
Diadino	0.55	0.76	37.10	0.20	0.74
Fuco	0.74	0.83	52.22	0.33	0.86
Hex19	0.51	0.64	34.15	0.15	0.71
Perid	0.49	0.72	34.77	0.14	0.70
Zea	0.35	0.72	35.78	0.15	0.58
PSC	0.82	0.54	31.61	0.16	0.90
PPC	0.68	0.53	32.91	0.15	0.82

The three model types each had different capabilities regarding pigment species retrieval. The Chl-a model showed a reasonable capacity for retrieving Chl-$c_{1,2}$, Chl-c, Diadino, Fuco, PSC, and PPC with satisfactory statistics. The Gaussian model could retrieve the four main pigment groups from phytoplankton absorption spectra with satisfactory results, except for Chl-b. In comparison, an obvious advantage was exhibited by the PLS model in retrieving more types of pigment that included Chl-a, Chl-b, Chl-$c_{1,2}$, Chl-c_3, Chl-c, Diadino, Fuco, Hex19, PSC, and PPC, even when there were no strong correlations between those pigments and Chl-a (Table 1).

Inversion accuracy is an important index with which to evaluate a model that can be illustrated intuitively using Taylor diagrams. Regardless of the model used, Chl-a (except for the Chl-a model), Chl-$c_{1,2}$ (except for the Gaussian model), Chl-c, PSC, and PPC were all retrieved well with relatively high accuracy. For But19, the Chl-a model and the PLS model produced similar results with lower retrieval precision, whereas the Gaussian model could not produce an estimate. In comparison with the Chl-a model, the capacity for retrieving Allo, Chl-c_3, and even Hex19, Perid, and Zea could be improved using the PLS model, as indicated by the higher r^2 values and lower RMSE, Bias, and MAPD values in the CV calculation (Table 7). The retrieval accuracy for Chl-b predicted by the PLS model

was higher than that of both the Chl-a model and the Gaussian model. The data points retrieved using the PLS model for these pigments were close to the reference point "A" shown in the Taylor diagrams (Figure 9), highlighting the improvement and advantages of this model. Among all the pigments, the best results for Chl-c, Chl-$c_{1,2}$, Diadino, and PSC were obtained by the Chl-a model with higher accuracy and lower error; Chl-a was retrieved with slightly higher accuracy by the Gaussian model.

4. Discussion

4.1. Performances of the Three Model Types

Using each of the three model types, stable and satisfactory results could be obtained for Chl-a, Chl-c, and PSC but not for Chl-b, which might reflect their different operating principles. Easy implementation is the main advantage of the Chl-a model, which is determined by the internal relationships between Chl-a and the objective pigments. Pigments that have a strong correlation with Chl-a, e.g., Chl-$c_{1,2}$, Chl-c, Diadino, Fuco, and PSC, could be retrieved with relatively high accuracy. The Gaussian model produced a better performance for Chl-a but it was limited with regard to the other specific accessory pigments owing to the similar absorption features within each pigment group, as noted by Bricaud et al. [37]. In comparison, the PLS model exhibited a considerable advantage for more accessory pigments by capturing the principal characteristics of phytoplankton absorption spectra. It means that the PLS model has the potential capability to obtain more types of phytoplankton community structure. This is important for the estimation of ocean productivity and quantitative evaluation of climate change. As reported in previous studies, it is difficult to retrieve Chl-b with satisfactory accuracy [7,9,24,38]. In this study, the Chl-a model also showed low inversion accuracy for Chl-b owing to the weak covariance between Chl-b and Chl-a. A similar result was observed using the Gaussian model, which was found to be related to the poor correlation between Chl-b and the absorption at a single band. However, a significant improvement for Chl-b retrieval was realized using the PLS model, which could reflect its use of the full spectral absorption signal. Better performance of the PLS model was also observed for other pigments, e.g., Allo, Chl-c_3, Hex19, Perid, and Zea.

For the examined models, it is important to discuss the internal variability of the dataset used. As reported by Sathyendranath et al. [26], the retrieval accuracy of Chl-b derived from an empirical regression model could be improved by restricting the samples to a lower ratio of Fuco/Chl-a. Such an improvement of r^2 for Chl-b was negligibly small based on our dataset, despite the repetition of the same analysis. A regionally developed model produced better performance than that obtained with a large-scale dataset, as revealed by comparison between the higher accuracy results derived by Liu et al. [9] and the relatively lower accuracy results obtained in our study, both of which were retrieved using the Gaussian model. Besides, results of Chl-$c_{1,2}$, Chl-b, and PPC obtained from the Chl-a model trained in our study were compared to that trained in Chase et al. [7], who built a global Chl-a model based on a global dataset of in situ hyperspectral reflectance. Similar results were obtained for Chl-$c_{1,2}$ and PPC in both studies, but an obvious difference existed for Chl-b. Phytoplankton adjusts their activities such as photosynthesis to adapt to specific and changeable conditions. Such complex and dynamic vital processes highlight the importance of regional models for analyzing phytoplankton community structures. Sample concentration also has an important influence on the model extraction of information regarding pigments. Differences in retrieval accuracy for PSC and PPC were found between our study and that of Liu et al. [9], i.e., the concentration of PPC was lower and that of PSC was approximately one order of magnitude lower than that in our dataset (Table 1). Lower concentrations of pigments might cause relatively weak signals, making pigment concentration inversion more difficult.

4.2. Influences of Derivation and Spectral Resolution on the PLS Model

Overall, $a_B(\lambda)$, $a''_B(\lambda)$, and $a_B^{4th}(\lambda)$ presented similar performances for most pigments, except Diadino, Fuco, and Perid. Using derivative spectra resulted in better performance for Diadino and poorer results for Fuco and Perid. Such results could be explained by the smooth and wide spectral features of Fuco and Perid, as shown in Figure 2 of Clementson and Wojtasiewicz [39]. In this case, a derivative analysis might alter the spectral signals. Conversely, Diadino showed relatively "sharp" absorption spectra within the blue wavebands, and derivative analysis could be beneficial for the extraction of "sharp" signals. Therefore, the superiority of derivative analysis depends on the features of the target pigment spectra.

The effect of spectral resolution on retrieval accuracy was evaluated by comparing continuous hyperspectral data with data collected from multispectral wavebands. Results showed that phytoplankton absorption spectra with hyperspectral resolution might exhibit considerable advantages for retrieval of certain individual pigments in comparison with the performance achieved using data with reduced spectral resolution. Our study suggests that the effectiveness of the multispectral model depends on whether the multispectral absorption contains the main spectral features with respect to the target pigment. However, the bands configured on current satellite sensors make it difficult to cover all spectral features of various pigments; therefore, it is undeniable that inversion performance could be improved using higher resolution spectra. This finding is consistent with the investigation of Wolanin et al. [40], which suggested that continuous hyperspectral data usually produce better results in comparison with multi-band data and that retrievals of pigment concentration could be improved by adding specific bands containing the optical signatures of those pigments.

5. Conclusions

We explored a number of methods for the retrieval of pigments concentrations from bio-optical measurements, including an empirical model based on chlorophyll-a, a Gaussian model, and a PLS regression model applied to phytoplankton absorption data after they were subjected to principal component analysis. According to validation of the model using in situ data, the Gaussian model could retrieve Chl-a, Chl-c, PSC, and PPC with satisfactory results. Chl-$c_{1,2}$, Chl-c, Diadino, Fuco, PSC, and PPC could be well derived using the Chl-a model because these pigments have strong covariance with the concentration of Chl-a. For pigments that have a poor correlation with Chl-a, the PLS model represents the preferable application prospect. According to our results, we offer the following recommendation. If the pigment has a strong and stable correlation with Chl-a, the use of the Chl-a model is recommended. If only the Chl-a concentration is to be retrieved, the use of the Gaussian model should be considered. The PLS model based on hyperspectral data is recommended for simultaneous inversion of multiple pigment concentrations.

Our study demonstrated the superiority of continuous hyperspectral phytoplankton absorption over the multispectral bands of existing sensors for retrieving pigment concentrations, at least for some auxiliary pigments. The result of this work is a step towards the development of satellite-based methods applicable to hyperspectral missions. To extend the results to remote sensing data, additional long-term measurements of in situ observations and further analysis of the relationships between R_{rs} and phytoplankton absorption will be necessary.

Author Contributions: Conceptualization, Y.Z., G.W. and S.S.; Methodology, Y.Z. and G.W.; Software, Y.Z., G.W., W.X., Y.X. and L.J.; Formal analysis, Y.Z. and G.W.; Supervision, G.W. and S.S.; Writing—original draft, Y.Z.; Writing—review and editing, Y.Z., G.W., S.S. and W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (41776045), the Simons Foundation grant Computational Biogeochemical Modeling of Marine Ecosystems (CBIOMES, 549947, S.S.), the Fundamental Research Funds for the Central Universities at Ho-

hai University (2017B06714, B200201013), and the Natural Science Foundation of Jiangsu Province (Grants No. BK 20200517).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the author upon reasonable request.

Acknowledgments: Additional support from the National Centre for Earth Observations (UK) is gratefully acknowledged. We thank James Buxton, MSc, from Liwen Bianji (Edanz) (www.liwenbianji.cn/ (accessed on 10 December 2021)), for editing the English text of a draft of this manuscript. Finally, the authors wish to thank the anonymous reviewers for their comments and feedback, which helped us to greatly improve the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Le Quéré, C.; Harrison, S.P.; Prentice, C.I.; Buitenhuis, E.T.; Aumonts, O.; Bopp, L.; Claustre, H.; Cunha, L.C.D.; Geider, R.; Giraud, X.; et al. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Glob. Chang. Biol.* **2005**, *11*, 2016–2040. [CrossRef]
2. Stocker, T.F.; Qin, D.; Plattner, G.-K.; Alexander, L.V.; Allen, S.K.; Bindoff, N.L. Technical summary—IPCC. In *Climate Change 2013: The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2013.
3. Sosik, H.M.; Sathyendranath, S.; Uitz, J.; Bouman, H.; Nair, A. In situ methods of measuring phytoplankton functional types. In *Phytoplankton Functional Types from Space*; Sathyendranath, S., Ed.; Reports of the International Ocean-Colour Coordinating Group (IOCCG), No. 15; IOCCG: Dartmouth, NS, Canada, 2014; pp. 21–38.
4. Bracher, A.; Bouman, H.; Brewin, B.; Wolanin, A. Obtaining Phytoplankton Diversity from Ocean Color: A Scientific Roadmap for Future Development. *Front. Mar. Sci.* **2017**, *4*, 55. [CrossRef]
5. Chase, A.; Boss, E.; Zaneveld, R.; Bricaud, A.; Claustre, H.; Ras, J.; Dall'Olmo, G.; Westberry, T.K. Decomposition of in situ particulate absorption spectra. *Limnol. Oceanogr.-Meth.* **2013**, *7*, 110–124. [CrossRef]
6. Bracher, A.; Taylor, M.H.; Taylor, B.; Dinter, T.; Röttgers, R.; Steinmetz, F. Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. *Ocean Sci.* **2015**, *11*, 139–158. [CrossRef]
7. Chase, A.; Boss, E.S.; Cetinić, I.; Slade, W.H. Estimation of Phytoplankton Accessory Pigments from Hyperspectral Reflectance Spectra: Toward a Global Algorithm. *J. Geophys. Res.* **2017**, *122*, 9725–9743. [CrossRef]
8. Catlett, D.; Siegel, D.A. Phytoplankton Pigment Communities Can be Modeled Using Unique Relationships with Spectral Absorption Signatures in a Dynamic Coastal Environment. *J. Geophys. Res.* **2018**, *123*, 246–264. [CrossRef]
9. Liu, Y.; Boss, E.; Chase, A.; Xi, H.; Zhang, X.; Röttgers, R.; Pan, Y.; Bracher, A. Retrieval of Phytoplankton Pigments from Underway Spectrophotometry in the Fram Strait. *Remote Sens.* **2019**, *11*, 318. [CrossRef]
10. Craig, S.E.; Jones, C.T.; Li, W.K.W.; Lazin, G.; Horne, E.; Caverhill, C.; Cullen, J.J. Deriving optical metrics of ecological variability from measurements of coastal ocean colour. *Remote Sens. Environ.* **2012**, *119*, 72–83. [CrossRef]
11. Niroumand-Jadidi, M.; Bovolo, F.; Bruzzone, L. Water Quality Retrieval from PRISMA Hyperspectral Images: First Experience in a Turbid Lake and Comparison with Sentinel-2. *Remote Sens.* **2020**, *12*, 3984. [CrossRef]
12. Giardino, C.; Bresciani, M.; Fabbretto, A.; Ghirardi, N.; Mangano, S.; Pellegrino, A.; Vaiciute, D.; Braga, F.; Brando, V.E.; Laanen, M.; et al. Hyperspectral Prisma Products of Aquatic Systems. In Proceedings of the IGARSS 2021-IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021. [CrossRef]
13. Casey, K.A.; Rousseaux, C.S.G.; Gregg, W.W.; Boss, E.S.; Chase, A.; Craig, S.; Mouw, C.B.; Reynolds, R.A.; Stramski, D.; Ackleson, S. A global compilation of in situ aquatic high spectral resolution inherent and apparent optical property data for remote sensing applications. *Earth Syst. Sci. Data* **2020**, *12*, 1123–1139. [CrossRef]
14. Isada, T.; Hirawake, T.; Kobayashi, T.; Nosaka, Y.; Natsuike, M.; Imai, I.; Suzuki, K.; Saitoh, S.I. Hyperspectral optical discrimination of phytoplankton community structure in Funka Bay and its implications for ocean color remote sensing of diatoms. *Remote Sens. Environ.* **2015**, *159*, 134–151. [CrossRef]
15. Xi, H.; Losa, S.N.; Mangin, A.; Garnesson, P.; Bretagnon, M.; Demaria, J.; Soppa, M.; Dandon, O.H.F.; Bracher, A. Global Chlorophyll a Concentrations of Phytoplankton Functional Types with Detailed Uncertainty Assessment Using Multisensor Ocean Color and Sea Surface Temperature Satellite Products. *J. Geophys. Res.* **2021**, *126*, e2020JC017127. [CrossRef]
16. Craig, S.; Lohrenz, S.E.; Lee, Z.; Mahoney, K.; Kirkpatrick, G.J.; Schofield, O.; Steward, R.G. Use of hyperspectral remote sensing reflectance for detection and assessment of the harmful alga, Karenia brevis. *Appl. Opt.* **2006**, *45*, 5414–5425. [CrossRef] [PubMed]
17. Lubac, B.; Loisel, H.; Guiselin, N.; Astoreca, R.; Artigas, L.F.; Meriaux, X. Hyperspectral and multispectral ocean color inversions to detect Phaeocystis globosa blooms in coastal waters. *J. Geophys. Res.* **2008**, *113*, 1–17. [CrossRef]

18. Uitz, J.; Stramski, D.; Reynolds, R.A.; Dubranna, J. Assessing phytoplankton community composition from hyperspectral measurements of phytoplankton absorption coefficient and remote-sensing reflectance in open-ocean environments. *Remote Sens. Environ.* **2015**, *171*, 58–74. [CrossRef]
19. Hirata, T.; Hardman-Mountford, N.J.; Brewin, R.J.W.; Aiken, J.; Barlow, R.; Suzuki, K.; Yamanaka, Y. Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences* **2011**, *8*, 311–327. [CrossRef]
20. Sun, D.; Pan, T.; Wang, S.; Hu, C. Linking phytoplankton absorption to community composition in Chinese marginal seas. *Prog. Oceanogr.* **2021**, *192*, 102517. [CrossRef]
21. Hoepffner, N.; Sathyendranath, S. Effect of pigment composition on absorption properties of phytoplankton. *Mar. Ecol. Prog. Ser.* **1991**, *73*, 11–23. [CrossRef]
22. Bidigare, R.R.; Morrow, J.; Kiefer, D.A. Derivative analysis of spectral absorption by photosynthetic pigments in the western Sargasso Sea. *J. Mar. Res.* **1989**, *47*, 323–341. [CrossRef]
23. Organelli, E.; Bricaud, A.; Antoine, D.; Uitz, J. Multivariate approach for the retrieval of phytoplankton size structure from measured light absorption spectra in the Mediterranean Sea (BOUSSOLE site). *Appl. Opt.* **2013**, *52*, 2257–2273. [CrossRef]
24. Hoepffner, N.; Sathyendranath, S. Determination of the major groups of phytoplankton pigments from the absorption spectra of total particulate matter. *J. Geophys. Res.* **1993**, *98*, 22789–22803. [CrossRef]
25. Ye, H.; Zhang, B.; Liao, X.; Li, T.; Shen, Q.; Zhang, F.; Zhu, J.; Li, J. Gaussian decomposition and component pigment spectral analysis of phytoplankton absorption spectra. *J. Oceanol. Limnol.* **2019**, *37*, 1542–1554. [CrossRef]
26. Sathyendranath, S.; Stuart, V.; Platt, T.; Bouman, H.; Ulloa, O.; Maass, H. Remote sensing of ocean colour:Towards algorithms for retrieval of pigment composition. *Indian J. Geo-Mar. Sci.* **2005**, *34*, 333–340.
27. Sathyendranath, S.; Stuart, V.; Irwin, B.D.; Maass, H.; Savidge, G.; Gilpin, L.; Platt, T. Seasonal variations in bio-optical properties of phytoplankton in the Arabian Sea. *Deep-Sea Res.* **1999**, *46*, 633–653. [CrossRef]
28. Stuart, V.; Sathyendranath, S.; Head, E.; Platt, T.; Maass, H. Bio-optical characteristics of diatom and prymnesiophyte populations in the Labrador Sea. *Mar. Ecol. Prog. Ser.* **2000**, *201*, 91–106. [CrossRef]
29. Stuart, V.; Ulloa, O.; Alarcón, G.; Sathyendranath, S.; Major, H.; Head, E.; Platt, T. Bio-optical characteris-tics of phytoplankton populations in the upwelling sys-tem off the coast of Chile. *Rev. Chil. Hist. Nat.* **2004**, *77*, 87–105. [CrossRef]
30. Hoepffner, N.; Sathyendranath, S. Bio-optical characteristics of coastal waters: Absorption spectra of phytoplankton and pigment distribution in the western North Atlantic. *Limnol. Oceanogr.* **1992**, *37*, 1660–1679. [CrossRef]
31. Kyewalyanga, M.; Platt, T.; Sathyendranath, S.; Lutz, V.A.; Stuart, V. Seasonal variations in physiological parameters of phytoplankton across the North Atlantic. *J. Plankton Res.* **1998**, *20*, 17–42. [CrossRef]
32. Head, E.J.H.; Horne, E. Pigment transformation and vertical flux in an area of convergence in the North Atlantic. *Deep-Sea Res.* **1993**, *40*, 329–346. [CrossRef]
33. Aiken, J.; Pradhan, Y.; Barlow, R.; Lavender, S.J.; Poulton, A.J.; Holligan, P.; Hardman-Mountford, N.J. Phytoplankton pigments and functional types in the Atlantic Ocean: A decadal assessment, 1995–2005. *Deep. Res. II Top. Stud. Oceanogr.* **2009**, *56*, 899–917. [CrossRef]
34. Bracher, A.; Xi, H.; Dinter, T.; Mangin, A.; Strass, V.; Appen, W.V.; Wiegmann, S. High Resolution Water Column Phytoplankton Composition Across the Atlantic Ocean from Ship-Towed Vertical Undulating Radiometry. *Front. Mar. Sci.* **2020**, *7*, 235. [CrossRef]
35. Xi, H.; Losa, S.N.; Mangin, A.; Soppa, M.; Garnesson, P.; Demaria, J.; Liu, Y.; Dandon, O.H.F.; Bracher, A. Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. *Remote Sens. Environ.* **2020**, *240*, 111704. [CrossRef]
36. Tsai, F.; Philpot, W. Derivative analysis of hyperspectral data. *Remote Sens. Environ.* **1998**, *66*, 41–51. [CrossRef]
37. Bricaud, A.; Claustre, H.; Ras, J.; Oubelkheir, K. Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *J. Geophys. Res.* **2004**, *109*, C11010. [CrossRef]
38. Lutz, V.A.; Sathyendranath, S.; Head, E.J.H. Absorption coefficient of phytoplankton: Regional variations in the North Atlantic. *Mar. Ecol. Prog. Ser.* **1996**, *135*, 197–213. [CrossRef]
39. Clementson, L.; Wojtasiewicz, B. Dataset on the absorption characteristics of extracted phytoplankton pigments. *Data Brief* **2019**, *24*, 103875. [CrossRef] [PubMed]
40. Wolanin, A.; Soppa, M.; Bracher, A. Investigation of Spectral Band Requirements for Improving Retrievals of Phytoplankton Functional Types. *Remote Sens.* **2016**, *8*, 871. [CrossRef]