

Article

Global Retrieval of Diatom Abundance Based on Phytoplankton Pigments and Satellite Data

Mariana A. Soppa^{1,*}, Takafumi Hirata^{2,3}, Brenner Silva⁴, Tilman Dinter^{1,5}, Ilka Peeken^{1,6}, Sonja Wiegmann¹ and Astrid Bracher^{1,5}

¹ Alfred-Wegener-Institute Helmholtz Center for Polar and Marine Research, Am Handelshafen 12, D-27570 Bremerhaven, Germany; E-Mails: Ilka.Peeken@awi.de (I.P.); Sonja.Wiegmann@awi.de (S.W.); Astrid.Bracher@awi.de (A.B.)

² Faculty of Environmental Earth Science, Hokkaido University, N10W5, Kita-Ku, Sapporo 060-0810, Japan; E-Mail: tahi@ees.hokudai.ac.jp

³ CREST, Japan Science and Technology Agency, N10W5, Kita-Ku, Sapporo 060-0810, Japan

⁴ Laboratory for Climatology and Remote Sensing, Philipps-Universität Marburg, Deutschhausstr. 12, D-35037 Marburg, Germany; E-Mail: silvab@staff.uni-marburg.de

⁵ Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, D-28359 Bremen, Germany; E-Mail: dinter@iup.physik.uni-bremen.de

⁶ MARUM-Center for Marine Environmental Sciences, Leobener Str., D-28359 Bremen, Germany

* Author to whom correspondence should be addressed; E-Mail: msoppa@awi.de; Tel.: +49-471-4831-1785; Fax: +49-471-4831-1149.

External Editors: Raphael M. Kudela, Nicolas Baghdadi and Prasad S. Thenkabail

Received: 12 June 2014; in revised form: 28 September 2014 / Accepted: 12 October 2014 /

Published: 22 October 2014

Abstract: Diatoms are the major marine primary producers on the global scale and, recently, several methods have been developed to retrieve their abundance or dominance from satellite remote sensing data. In this work, we highlight the importance of the Southern Ocean (SO) in developing a global algorithm for diatom using an Abundance Based Approach (ABA). A large global *in situ* data set of phytoplankton pigments was compiled, particularly with more samples collected in the SO. We revised the ABA to take account of the information on the penetration depth (Z_{pd}) and to improve the relationship between diatoms and total chlorophyll-a (TChla). The results showed that there is a distinct relationship between diatoms and TChla in the SO, and a new global model ($ABA_{Z_{pd}}$) improved the estimation of diatoms abundance by 28% in the SO compared with the

original ABA model. In addition, we developed a regional model for the SO which further improved the retrieval of diatoms by 17% compared with the global ABA_{Zpd} model. As a result, we found that diatom may be more abundant in the SO than previously thought. Linear trend analysis of diatom abundance using the regional model for the SO showed that there are statistically significant trends, both increasing and decreasing, in diatom abundance over the past eleven years in the region.

Keywords: ocean colour; phytoplankton functional types; diatom; remote sensing; chlorophyll-a

1. Introduction

Phytoplankton is the basis of the marine food web and a key component in the marine ecosystem. The term phytoplankton functional type (PFT, see Table 1 for a list of abbreviations and symbols) is used to distinguish the different roles of the phytoplankton in the biogeochemical cycle of the oceans [1]. Of special interest are the diatoms, which are the major contributors to the oceanic primary production [2] and carbon export [3] and, together with dinoflagellates, the most diverse PFT [4,5]. Diatoms are also one of the largest PFTs in terms of size, ranging from micrometers to a few millimeters [4]. They tend to dominate the phytoplankton community in coastal, polar and upwelling regions, where waters are typically rich in nutrients. In the Southern Ocean (SO), they represent 89% of the primary production [2] and are found in high concentrations in stratified waters near ice edge-zones [6,7], but also frequently form blooms at the Polar Front [8].

Given the biogeochemical and ecological importance of diatoms, it is necessary to understand how they respond to climate variability on global and regional scales, a task that cannot be achieved without knowledge of their temporal and spatial distribution. In the last decade, considerable effort has been invested in developing and improving approaches to retrieve the global distribution of diatoms from satellite data. Examples of these are PHYSAT [9], PhytoDOAS [10] and the Abundance Based Approach (ABA) [1,11–13]. PHYSAT determines dominance of diatoms (in addition to nano-eukaryotes, *Prochlorococcus*, *Synechococcus*-like and *Phaeocystis*-like) by identifying their specific spectral signatures from the normalized water-leaving radiance. Like PHYSAT, PhytoDOAS is based on analyzing optical (hyperspectral) information of satellite data and retrieves diatoms (as well as cyanobacteria, dinoflagellates and coccolithophores) by identifying their specific absorption in the backscattered solar radiation. The ABA, such as that by Hirata *et al.* [1], in contrast, is an ecological approach which applies satellite-measured chlorophyll-a (Chla) to empirical relationships between TChla and diatoms (as well as dinoflagellates, green algae, haptophytes, prokaryotes, pico-eukaryotes and *Prochlorococcus* sp.) derived from *in situ* measurements. However, unlike PHYSAT, both PhytoDOAS and ABA provide a quantitative estimation of the diatom abundance instead of its dominance.

Compared to optically based approaches, a great advantage of the ABA is the smaller computational effort; even if the satellite data volume becomes larger with higher temporal and spatial resolutions, the data processing load is not heavy and re-processing can also be done relatively easily.

The ABA can be applied to global level-2 or level-3 products of TChla, which are freely available to the scientific community, as opposed to, for example, PhytoDOAS method that uses the top of atmosphere radiance data (*i.e.*, level-1 product). In addition, the ABA has better global coverage of *in situ* data for model development and validation. Because ABA is an empirical model, its refinement is needed in the light of additional data to improve the retrieval of diatoms for both global and under-sampled oceans.

This paper focus on the retrieval of diatom abundance of Hirata *et al.* [1] based on two premises: (i) Diatoms are the major primary producers in the SO [2] and (ii) 90% of the diffusely reflected irradiance measured by ocean color sensors originates from the first optical depth, also referred to as the penetration depth (Z_{pd}) [14]. These premises are also general limitations of the existing ABA. Although Hirata *et al.* [1] used a large global data set of phytoplankton pigments, new measurements, particularly in the SO (defined here as the region south of 50°S), have become available since then. Main objectives of this paper are:

(1). Compilation of a new and larger global data set of *in situ* phytoplankton pigment profiles, including more measurements in the SO (Figure 1) which was not well covered previously, and to investigate the relationship between fractional contribution of diatoms and TChla using the new data set in comparison to previous findings.

(2). Refinement of the ABA to account for the pigment information in the Z_{pd} ($ABA_{Z_{pd}}$). In ABA [1], the fractional contribution of diatoms to TChla was estimated based on the previous work of Uitz *et al.* [15], who used the phytoplankton pigment concentration integrated over the euphotic depth (Z_{eu}). However, the pigment concentration estimated by the satellite sensor is an optically-weighted concentration in the Z_{pd} , which is approximately 4.6 times shallower than the Z_{eu} [16].

(3). Evaluation of the performance of the ABA (*i.e.*, $ABA_{Z_{pd}}$) for global oceans and for the SO region.

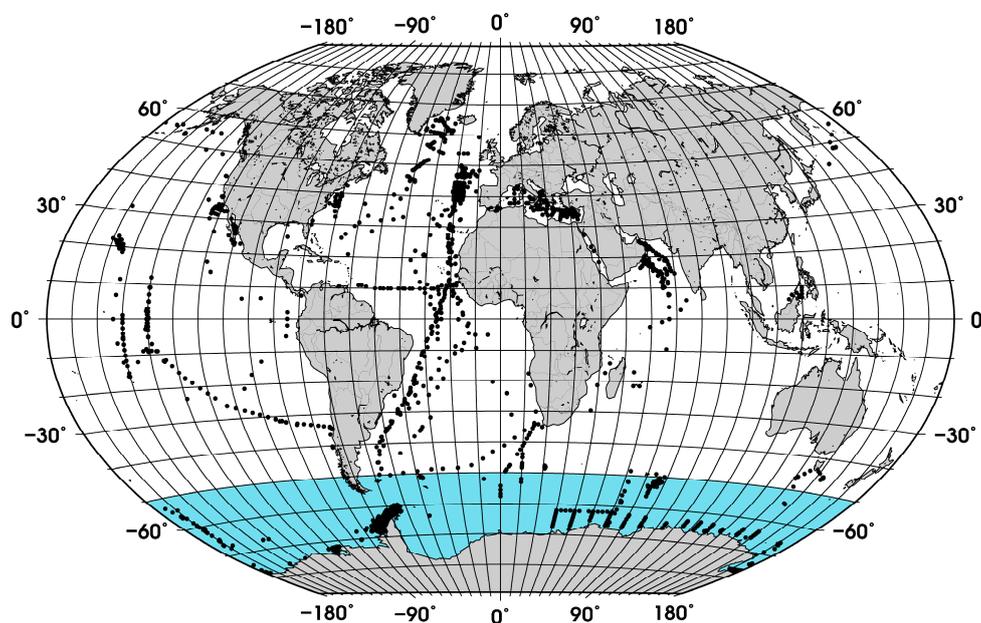
Table 1. Abbreviations and symbols used in this study and their meaning.

Abbreviations and Symbols	Meaning
a	partial coefficients from multiple regression analysis
ABA	Abundance Based Approach
$ABA_{Z_{pd}}$	modified ABA
ABA*	original model of Hirata <i>et al.</i> [1] parameterized with the $DP_{Z_{pd}}$ data set
ABA**	original model and fitting parameters of Hirata <i>et al.</i> [1]
Allo	alloxanthin, in mg/m^3
Butfuco	19'-butanoyloxyfucoxanthin, in mg/m^3
c	vector containing the partial coefficients corresponding to each $DP_{Z_{pd}}$ on log scale
Chla	chlorophyll-a concentration, in mg/m^3
Diatom $_{Z_{pd}}$	concentration of TChla $_{Z_{pd}}$ that is attributed to diatoms, in mg/m^3
DP	concentration of the diagnostic pigments, in mg/m^3
$DP_{Z_{pd}}$	DP weighted in the Z_{pd} , in mg/m^3
DP_w	estimated TChla, in mg/m^3
f	Fraction
f -Diatom	fraction of TChla attributed to diatoms
f -Diatom $_{Z_{pd}}$	same as f -Diatom, but calculated from $DP_{Z_{pd}}$

Table 1. Cont.

Abbreviations and Symbols	Meaning
f -PFT	fraction of TChla attributed to a phytoplankton functional type
Fuco	fucoxanthin, in mg/m^3
Hexfuco	19'-hexanoyloxyfucoxanthin, in mg/m^3
HPLC	High-Performance Liquid Chromatography
$M_{Z_{pd}}$	matrix containing the seven $DP_{Z_{pd}}$
MODIS	Moderate Resolution Imaging Spectroradiometer
Perid	peridinin, in mg/m^3
PFT	Phytoplankton Functional Type
SO	Southern Ocean, in this study defined as the region south of 50°S
TACC	sum of the concentration of all accessory pigments, in mg/m^3
$TACC_{Z_{pd}}$	TACC weighted in the Z_{pd} , in mg/m^3
TChla	sum of the concentration of monovinyl Chla, divinyl Chla, Chla allomers, Chla epimers and chlorophyllidae, in mg/m^3
$TChla_{Z_{pd}}$	TChla weighted in the Z_{pd} , in mg/m^3
TChlb	sum of monovinyl chlorophyll-b and divinyl chlorophyll-b, in mg/m^3
Z_{eu}	euphotic depth, given in m
Z_{pd}	penetration depth, given in m
Zea	zeaxanthin, in mg/m^3

Figure 1. Distribution of the quality controlled *in situ* measurements. The SO, region south of 50°S , is the portion of the global ocean presented in blue.



The Theoretical Basis of the ABA

The ABA calculates the fraction (f) of Chla attributed to a specific PFT (f -PFT) using concentrations of diagnostic pigments of phytoplankton (*i.e.*, Diagnostic Pigment Analysis–DPA, [15,17]). According to Uitz *et al.* [15], the Chla can be expressed by the sum of seven diagnostic pigments as:

$$DP_w = \sum_{i=1}^7 a_i DP_i \quad (1)$$

where DP_w is the estimated Chla, a are the partial coefficients and DP are the concentration of the diagnostic pigments: Fucoxantin (Fuco), peridinin (Perid), 19'-hexanoyloxyfucoxanthin (Hexfuco), 19'-butanoyloxyfucoxanthin (Butfuco), alloxanthin (Allo), monovinyl chlorophyll-b + divinyl chlorophyll-b (TChlb) and zeaxanthin (Zea). These diagnostic pigments represent the main PFTs (Table 2).

For example, the fraction of Chla that is attributed to diatoms (f -Diatom) is derived as:

$$f\text{-Diatom} = a_1 \text{Fuco} / DP_w \quad (2)$$

f -Diatom values lower than 0 and greater than 1 are set to 0 and 1, respectively.

Table 2. Phytoplankton pigments and their respective PFT (adapted from Vidussi *et al.* [17]).

Phytoplankton Pigments	PFT
Fuco	Diatoms
Perid	Dinoflagellates
Hexfuco	Prymensiophytes and chrysophytes
Butfuco	Prymensiophytes and chrysophytes
Allo	Cryptophytes
TChlb	Green algae and prochlorophytes
Zea	Cyanobacteria and prochlorophytes

Once the f -PFT has been determined, the relationship between f -PFT and Chla can be represented by a model or fit function and quantified, where the relationship varies according to the PFT. The model for f -Diatoms was previously shown [1] as:

$$f\text{-Diatom} = [a_0 + \exp(a_1 x + a_2)]^{-1} \quad (3)$$

where x represents $\log_{10}(\text{Chla})$ and a_0 , a_1 , a_2 are the fitting parameters with values of 1.3272, -3.9828 and 0.1953, respectively.

With the knowledge of the fit function, its parameters and Chla, it is possible to retrieve the f -Diatom, once Chla, which is operationally produced as a satellite product, is known. To retrieve the diatom abundance in terms of Chla (mg/m^3), the f -Diatom is multiplied by the Chla value of each sample.

2. Data and Methods

2.1. In Situ Measurements of Phytoplankton Pigments

A data set of phytoplankton pigment profiles measured with the High-Performance Liquid Chromatography (HPLC) technique was supplemented with data obtained from the SeaWiFS Bio-optical Archive and Storage System (SeaBASS, [18]), Marine Ecosystem Data (MAREDAT, [19]), and from the individual cruises KEOPS ([20]), Bonus Good Hope, ANT-XVIII/2 (EisenEx), ANTXXI/3 (EIFEX, [3]), ANT XXVI/3, ANT XXVIII/3, Sonne SO218 [21], Merian 18-3,

Meteor 55 and Meteor 60. The pigments from the cruises Meteor 55, Meteor 60, ANT XXVI/3 and ANT-XVIII/2 were measured in accordance with the method described in Hoffmann *et al.* [22] and for the cruises Merian 18-3 and ANT XXVIII/3 in accordance with that in Taylor *et al.* [23].

The data were quality controlled in a way similar to the method used by Uitz *et al.* [15] and Peloquin *et al.* [19]: (i) Samples with accessory pigment concentrations below 0.001 mg/m^3 were set to zero, (ii) samples with TChla below 0.001 mg/m^3 and fewer than 4 accessory pigments were excluded. The TChla was defined as the sum of monovinyl Chla, divinyl Chla, Chla allomers, Chla epimers and chlorophyllidae. To ensure that the profiles had a minimum vertical resolution, we restricted the data set to profiles with at least (i) one sample at the surface (0 to 12 m), (ii) one sample below the surface, (iii) samples collected at four or more different depths, and (iv) with one sample within the Z_{pd} . The last quality control measure was based on the \log_{10} -linear relationship between $TChla_{Z_{pd}}$ and the sum of all accessory pigments in the Z_{pd} ($TACC_{Z_{pd}}$). Data that fell outside the 95% confidence interval were removed.

In addition, samples located in coastal waters (<200 m) were excluded using the ETOPO1 bathymetry [24]. The final data set contained 3988 samples, which were randomly split into work (~70% of the data) and validation (~30% of the data) subsets (Figure 1). While the whole data set was used to calculate the partial coefficients used for estimating $f\text{-Diatom}_{Z_{pd}}$, the work and validation subsets were used for model development and validation of the $ABA_{Z_{pd}}$, respectively.

2.2. Satellite Data

Eleven years (2003–2013) of MODIS Aqua Level 3 4 km binned Chla data (R2013.0) were used. MODIS is a multispectral sensor on board of the Aqua satellite and with global coverage. The data were obtained from the OceanColor website [25] at daily temporal resolution. Monthly averages of diatom abundance were calculated onto a 10 minute grid and used to derive climatological maps of diatom abundance. To avoid coastal waters, where the retrieval of the ABA was not intended, we removed grid cells located in waters shallower than 200 m using the ETOPO1 bathymetry [24].

2.3. An Improved Abundance Based Approach

In previous approach [1], the $f\text{-Diatom}$ was calculated using the coefficients of Uitz *et al.* [15], which take account of the phytoplankton pigment integrated over the Z_{eu} . Here, we extended the ABA to take account of the information in the Z_{pd} . For this purpose, we recalculated the coefficients a (Equation (1)) using an updated global data set of HPLC phytoplankton pigment profiles ($N = 3988$). The weighted pigment concentration in the Z_{pd} ($DP_{Z_{pd}}$) was calculated as described in Gordon & Clark [26], with the diffusive attenuation coefficient at 490 nm derived from the profiles of TChla as described in Morel & Maritorena [27]. The Z_{pd} was computed as $Z_{pd} = Z_{eu}/4.6$, and Z_{eu} was derived from the surface TChla as $Z_{eu} = 34 \times TChla^{-0.39}$ (Morel, in Lee *et al.* [28]). Profiles were interpolated with 1-m increments from the deepest sample to the sample closest to the surface before the calculation of $DP_{Z_{pd}}$.

A limitation of retrieving PFTs from HPLC pigments is the presence of a DP in more than one PFT [1,15]. The quality controlled data set was corrected for Fuco to account for its co-existence in other PFTs, in accordance with Hirata *et al.* [1].

Nonlinear minimization was used to retrieve the partial coefficients, which represent the estimates of the TChla to the DP ratios [15]. The function to be minimized is expressed as:

$$\| \exp(c) M_{Z_{pd}} \| \rightarrow \min \quad (4)$$

where c is a vector containing the seven coefficients which each correspond to each $DP_{Z_{pd}}$ on the log scale, and $M_{Z_{pd}}$ a matrix containing the seven $DP_{Z_{pd}}$. The nonlinear minimization method requires an initial guess of c , which was obtained from the multiple linear regression analysis. The standard deviation of the coefficients is given by the square root of the diagonal elements of the inverse of the Hessian matrix.

Using the new coefficients, the f -Diatom $_{Z_{pd}}$ was calculated for each sample of the work and validation subsets. The work subset was then sorted according to the TChla $_{Z_{pd}}$ and smoothed with a 5-point running mean filter to improve the signal-to-noise ratio [1,12]. Next, the relationship between f -Diatom $_{Z_{pd}}$ and TChla $_{Z_{pd}}$ was quantified using a nonlinear least-square fit applied to the work subset and represented by a model and its fitting parameters. Once the model has been defined, satellite-derived TChla data was applied to the model to obtain the global distribution of f -Diatom $_{Z_{pd}}$. Diatom abundance (Diatom $_{Z_{pd}}$, mg/m³) is then obtained by multiplying f -Diatom $_{Z_{pd}}$ by TChla $_{Z_{pd}}$.

The accuracy of the new model was tested using the validation subset. The uncertainties were estimated by the mean absolute error (MAE, [29]) and maximum absolute error (Max. Abs. Error) between the modeled and the measured (*in situ*) Diatom $_{Z_{pd}}$. The models were compared by the difference between the MAE of the original model and the new model, relative to the original model, and expressed in percent (%). The data were log transformed prior to the calculation of the validation statistics. We used $\log_{10}(\text{data} + \lambda)$ where $\lambda = 0.00003$, approximately one half of the smallest non-zero value of the *in situ* Diatom $_{Z_{pd}}$ validation data, since the data set contained zeroes. In addition, to investigate whether using different partial coefficients results in significant changes in f -Diatom, we estimated f -Diatom using the coefficients of Uitz *et al.* [15] and Brewin *et al.* [30] and compared the results based on the coefficient of determination.

A Regional Model for the SO

The main difference between the SO model and the global model is that the relationship between Diatom $_{Z_{pd}}$ and TChla $_{Z_{pd}}$ is investigated not in terms of f -Diatom $_{Z_{pd}}$, but instead in terms of the concentration of TChla $_{Z_{pd}}$ that is attributed to diatoms, similar to the approach adopted by Brewin *et al.* [13] to retrieve phytoplankton size classes. As in Brewin *et al.* [13], the fit function was applied to \log_{10} -transformed data. To develop the regional model for the SO, we selected the samples of the global work and validation data sets that were located in the SO, creating a SO work and a validation data set with 1069 and 460 samples, respectively. The relationship between Diatom $_{Z_{pd}}$ and TChla $_{Z_{pd}}$ was investigated and validated. Note that for the work data set we applied the running mean exclusively to the SO data.

2.4. Statistical Analysis of Trends

Linear trends were computed for February from monthly standardized anomalies over the 2003–2013 period in the SO using the regional model. To remove the seasonal cycle we calculated the monthly

anomalies in diatom abundance for each grid cell by subtracting the climatological mean from the corresponding monthly mean (e.g., February 2003–climatology of February). The monthly anomalies were divided by the corresponding climatological standard deviation (e.g., standard deviation of February) to enable the direct comparison of trends between different regions (grid cells). The trends were computed using the non-parametric Kendall’s tau test with Sen’s method at the 95% confidence level and in grid cells with 100% temporal coverage.

3. Results and Discussion

3.1. The $ABA_{Z_{pd}}$

Table 3 shows the partial regression coefficients, and their respective standard deviation, calculated with Equation (3). For comparison, we also present the partial coefficients estimated by Uitz *et al.* [15], Brewin *et al.* [30] and Fujiwara *et al.* [31]. Comparing our coefficients with those from Uitz *et al.* [15], there is a notable difference, except for the coefficients of Fuco and TChlb. These differences result from the inclusion of more profiles, their geographical distribution, the adjustment of Fuco prior to the DPA analysis, and because we used the pigment concentration weighted in the Z_{pd} , while Uitz *et al.* [15] integrated the pigments over Z_{eu} . When compared to the two other studies, where the partial coefficients were derived from surface measurements, our coefficients are more similar to those described in Brewin *et al.* [30]. Brewin *et al.* [30] included measurements of five Atlantic Meridional Transect (AMT) cruises in the Atlantic Ocean, while Fujiwara *et al.* [31] used measurements from three cruises in the Western Arctic Ocean. Although our data set includes measurements from these regions, the number of samples in the Arctic region is fewer than that from the Atlantic (Figure 1).

Table 3. The partial regression coefficients and standard deviation (in brackets) where available. The number of samples is indicated by N. The empty fields indicate that the coefficient is not statistically significant.

Coefficients	Ocean	N	Fuco	Perid	Hexfuco	Butfuco	Allo	TChlb	Zea
Present study	Global	3988	1.554 (0.010)	0.413 (0.568)	0.855 (0.068)	1.174 (0.145)	2.387 (0.099)	1.062 (0.070)	2.037 (0.040)
Uitz <i>et al.</i> [15]	Global	2419	1.41 (0.02)	1.41 (0.10)	1.27 (0.02)	0.35 (0.15)	0.60 (0.16)	1.01 (0.10)	0.86 (0.09)
Brewin <i>et al.</i> [30]	Atlantic	466	1.72	1.27	0.68	1.42	4.96	0.81	1.28
Fujiwara <i>et al.</i> [31] *	Arctic	76	1.85	1.49	1.74		5.88	1.31	3.54

* standard errors are less than 1.

Moreover, we have re-run the analysis taking into consideration the surface samples (<12 m) from our profiles and observed only a slight difference in the coefficient of Fuco (1.531) as compared to the weighted Z_{pd} concentrations. Except for Perid and Hexfuco, the standard deviation of our coefficients are much lower than, or similar to, the ones obtained by Uitz *et al.* [15].

Nonetheless, we observed very similar *f*-Diatom values when using the partial coefficients of Uitz *et al.* [15], Brewin *et al.* [30] and ours. The coefficients of determination are higher than 0.98, suggesting the choice of partial coefficients has no influence on the retrievals of *f*-Diatom, which is

consistent with Brewin *et al.* [30]. Brewin *et al.* [30] compared size-fractionated chlorophyll (SFC) estimated from phytoplankton pigment data and calculated using Uitz *et al.* [15] partial coefficients and their own, with size-fractionated filtration (SFF) measurements. They observed biases between SFC and SFF for nanoplankton and picoplankton size classes; however, the variations in the partial coefficients did not influence the results significantly. The high correlation between the $TChla_{Zpd}$ and DP_w , with DP_w calculated using Equation (4) ($r^2 = 0.85$, $DP_w = 0.86 TChla_{Zpd} + 0.074$, $N = 3988$, $p < 0.001$), gives us confidence to use the partial coefficients to determine the f -Diatom.

Figure 2 shows the change in the f -Diatom $_{Zpd}$ with increasing $TChla_{Zpd}$. The green and blue lines represent the new model (ABA_{Zpd}) and the model of Hirata *et al.* [1] (ABA^*), respectively, parameterized with the DP_{Zpd} data set. The red line represents the original model and fitting parameters of Hirata *et al.* [1] (ABA^{**}). It can be seen that diatoms are dominant at high $TChla_{Zpd}$ (Figures 2a,b), which is consistent with previous studies [1] even if a significant number of new samples were added in our dataset. Moreover, we also observed unusually high f -Diatom $_{Zpd}$ in low $TChla_{Zpd}$ waters ($<0.1 \text{ mg/m}^3$, $N = 670$). Taking a closer look at the profiles, in which $Fuco_{Zpd}$ corresponded to at least 50% of the $TACC_{Zpd}$, we observed that most of the data sets (12 out of 16) are from samples taken in Antarctic, in the East Antarctic marginal ice zone (BROKE cruise, [6]). On average, the ratio of $Fuco_{Zpd}$ to $TChla_{Zpd}$ is 0.165 for the entire DP_{Zpd} data set, 0.071 excluding the SO data, but 0.317 for the SO data, indicating higher f -Diatom $_{Zpd}$ values in low $TChla_{Zpd}$ waters in the SO.

Figure 2. Relationship between $TChla_{Zpd}$ and f -Diatom $_{Zpd}$: (a) Global data set ($N = 2806$), (b) global data set excluding SO data ($N = 1737$) and (c) SO data ($N = 1069$). The data sets were smoothed with a 5 point running mean to improve the signal-to-noise ratio [1]. The green and blue lines represent the new model (ABA_{Zpd}) and the model of Hirata *et al.* [1] (ABA^*) parameterized with the DP_{Zpd} data set. The red line represents the original model and fitting parameters of Hirata *et al.* [1] (ABA^{**}). The fitting parameters are presented in Table 4. The MAE values refer to the errors in terms of f -Diatom $_{Zpd}$. Note that we could not fit the global models to the SO data set exclusively. The cyan and green lines in (c) represent the regional model for the SO and the ABA_{Zpd} plotted with the global fitting parameters as reference.

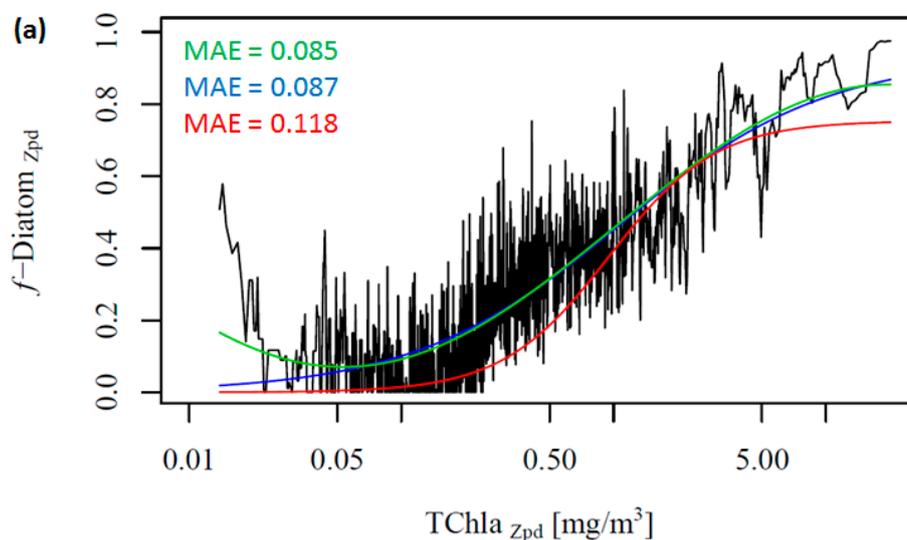


Figure 2. Cont.

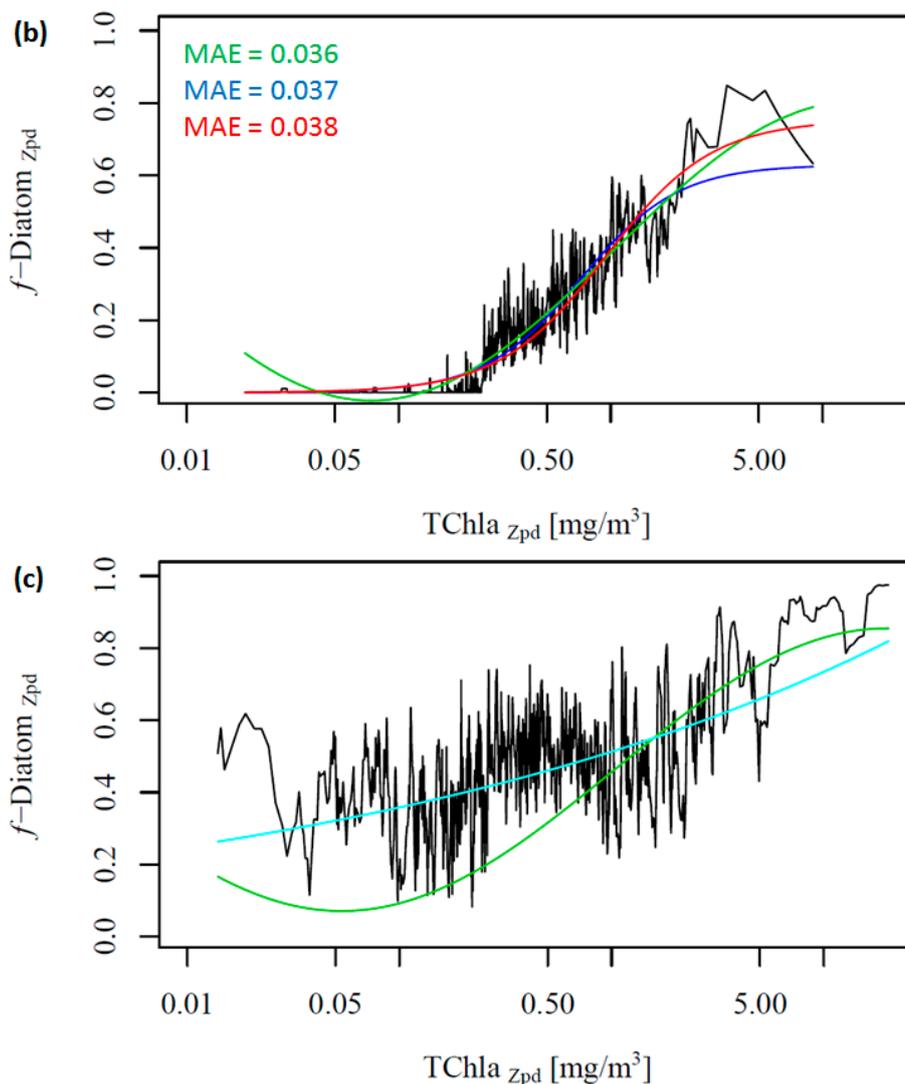


Table 4. Models of $f\text{-Diatom}_{z_{pd}}$ as a function of $\text{TChla}_{z_{pd}}$ and their respective fitting parameters used to plot the lines in Figure 2a,b. Note that we could not fit the global models to the SO data exclusively. The fitting parameters of the original ABA model of Hirata *et al.* [1] (ABA**) do not change and therefore they are presented only once in the table.

	$f\text{-Diatom}_{z_{pd}}$	Model	a_0	a_1	a_2	a_3
Global data set	$\text{ABA}_{z_{pd}}$	$a_0 + a_1 \sin(a_2(x + a_3))$	0.4629	0.3921	1.2214	-0.01412
	ABA*	$[a_0 + \exp(a_1x + a_2)]^{-1}$	1.0733	-2.0484	0.1314	-
	ABA**	$[a_0 + \exp(a_1x + a_2)]^{-1}$	1.3272	-3.9828	0.1953	-
Global data set excluding SO data	$\text{ABA}_{z_{pd}}$	$a_0 + a_1 \sin(a_2(x + a_3))$	0.3909	0.4131	1.3763	-0.0114
	ABA*	$[a_0 + \exp(a_1x + a_2)]^{-1}$	1.5890	-4.3778	-0.1521	-

$x = \log_{10}(\text{TChla})$; * model of Hirata *et al.* [1] parameterized with the $\text{DP}_{z_{pd}}$ data set; ** original model and fitting parameters of Hirata *et al.* [1].

Considering our newer and larger data set, Hirata *et al.* [1] considerably underestimates f -Diatom_{Zpd} in almost the entire TChla_{Zpd} range (Figure 2a-red line). This is partly due to the difference in the data set used. When we fit their model to the new data set, the model is found to fit well to the data, as indicated by the low errors (Table 5 and Figure 2a-blue line). However, it fails when predicting f -Diatom_{Zpd} in very low TChla_{Zpd} waters, mostly for the SO. Thus, we test a new model, a sinusoidal function to better fit this observed trend in the SO (Table 4, Figure 2-green line). The ABA_{Zpd} and ABA* produce almost identical curves for TChla_{Zpd} above 0.065 mg/m³ and similar fitting and validation statistics. The ABA** model provides accurate retrieval of diatoms globally. However, it produces larger errors than the ABA_{Zpd} and the ABA* do for the SO. The ABA_{Zpd} improves the MAE by 27.96% for the SO (Table 6 and Figure S1).

Table 5. Statistical results of the fits for the global data set and global excluding SO data using the fitting parameters of Table 4. Note that we could not fit the global models to the SO data exclusively. The fitting statistics for the SO data set refer to the regional SO model (Figure S2 in the Supplementary Information). The MAE and Max. Abs. Error are given in f -Diatom_{Zpd} for the global models and for the regional model in mg/m³ (log₁₀-transformed data).

(a) Fit		N	r ²	p-value	MAE	Max. Abs. Error
Global data set	ABA _{Zpd}	2806	0.71	=0	0.085	0.474
	ABA*	2806	0.70	=0	0.087	0.559
	ABA**	2806	0.66	=0	0.118	0.609
Global data set excluding SO data	ABA _{Zpd}	1737	0.89	<0.001	0.036	0.216
	ABA*	1737	0.88	<0.001	0.037	0.249
	ABA**	1737	0.88	<0.001	0.038	0.246
SO data set	Regional model	1069	0.95	<0.001	0.104	0.695

* model of Hirata *et al.* [1] parameterized with the DP_{Zpd} data set; ** original model and fitting parameters of Hirata *et al.* [1].

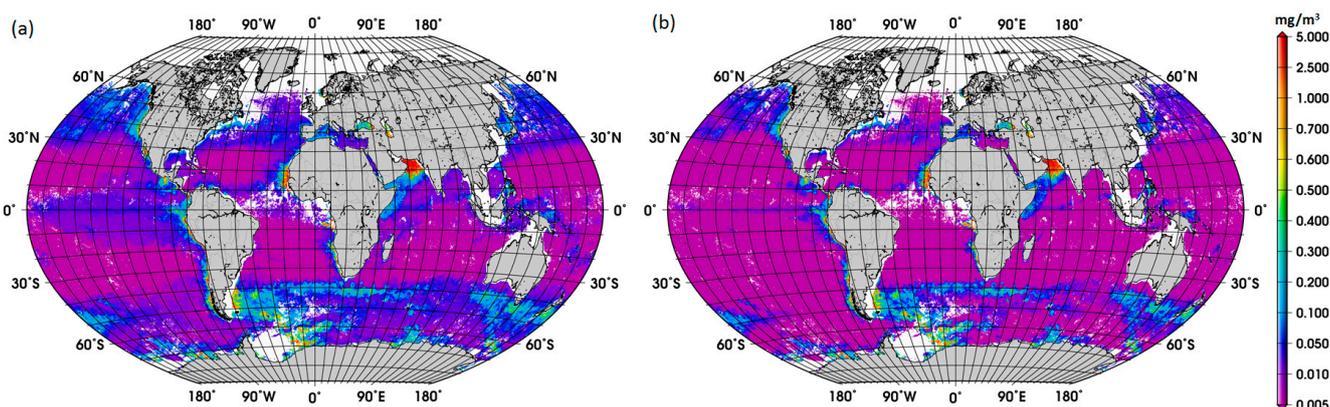
Table 6. Statistical results of the validation in terms of diatoms abundance. Note that we could not fit the global models to the SO data exclusively. The results for the SO data set correspond to the global models using the global fitting parameters and the regional model. The MAE and Max. Abs. Error are given in mg/m³. The statistics were calculated with log₁₀-transformed data (e.g., log₁₀(y + 0.00003)).

Validation		N	r ²	p-value	MAE	Max. Abs. Error
Global data set	ABA _{Zpd}	1182	0.57	<0.001	1.219	4.048
	ABA*	1182	0.55	<0.001	1.217	4.044
	ABA**	1182	0.57	<0.001	1.035	3.945
Global data set excluding SO data	ABA _{Zpd}	722	0.59	<0.001	0.883	3.160
	ABA*	722	0.68	<0.001	1.195	3.068
	ABA**	722	0.69	<0.001	1.200	3.024
SO data set	ABA _{Zpd}	460	0.40	<0.001	0.559	4.048
	ABA*	460	0.39	<0.001	0.562	4.044
	ABA**	460	0.39	<0.001	0.776	3.945
	Regional model	460	0.39	<0.001	0.465	4.128

* model of Hirata *et al.* [1] parameterized with the DP_{Zpd} data set; ** original model and fitting parameters of Hirata *et al.* [1].

To further investigate the influence of the SO data, we removed these data from the work data set (38% of the data), recalculated f -Diatom_{Z_{pd}}, and fitted the models (Figure 2b and Table 4). The comparison of Figure 2a,b shows clearly the influence of the SO data, which is responsible for most of the data spread in Figure 2a as well as for the high f -Diatom_{Z_{pd}} in low TChla_{Z_{pd}} waters. When we exclude the SO data from the analysis, the fits improve greatly the MAE decrease to values close to 0.04 (Table 5). In addition, it leads to a better representation of the diatom abundance in oligotrophic waters, as well as to an underestimation of the actual f -Diatom_{Z_{pd}} in the SO, as shown in Figure 3. The advantage of including the SO data is a more realistic retrieval of diatoms in the SO, but an overestimation in other regions of low TChla_{Z_{pd}}. While the *in situ* data show that the f -Diatom_{Z_{pd}} might be very low (~ 0) at very low TChla_{Z_{pd}} (e.g., in oligotrophic gyres), the predicted f -Diatom_{Z_{pd}} presents values higher than zero, overestimating f -Diatom_{Z_{pd}} in the oligotrophic gyres.

Figure 3. Monthly mean TChla_{Z_{pd}} (mg/m³) of diatoms for February 2003 using the ABA_{Z_{pd}} model parameterized with: (a) Global data set (average = 0.060 mg/m³) and (b) global data set excluding SO data (average = 0.041 mg/m³). White areas correspond to waters with depths shallower than 200 m or without satellite information.



It should be noted that the model used to retrieve f -Diatom_{Z_{pd}} as a function of TChla_{Z_{pd}} was empirically built upon *in situ* data sets, which showed that diatoms tend to be the dominant PFT at high TChla. However, this may not be the case of blooms of mixed PFTs occur, or with a different PFT (e.g., coccolithophores close to New Zealand [32]), as pointed out by Brewin *et al.* [13]. In such cases, additional information on PFTs derived from methods that do not depend on this assumption (e.g., PhytoDOAS) may improve the knowledge on the diatom abundance and the distribution pattern.

Moreover, we did not obtain significant results in fitting the two global models to the SO data exclusively (Figure 2c, ABA_{Z_{pd}} plotted as reference). The diatoms in the SO exhibit a variability which is different from other oceanic regions (e.g., the North Pacific and the North Atlantic), and there is a need for a regional SO model. Thus, we developed a regional model for the SO, and the relationship between TChla_{Z_{pd}} and Diatom_{Z_{pd}} in the SO can be expressed as: $\log_{10}(y) = 1.1559\log_{10}(x) + (-0.2901)$ (Figure S2 in the Supplementary Information). The validation results of the SO model show that the regional model is consistent and more appropriate than the global ABA_{Z_{pd}} model for retrieving diatoms in the SO (Table 6 and Figure S2 in the Supplementary Information). The regional model improved by 17% the retrieval of diatoms abundance in the SO compared with the ABA_{Z_{pd}}.

The ideal global retrieval of diatoms should apply the ABA_{Zpd} model parameterized with the global data set excluding SO data (Figure 2b green line) to the region north of 50°S, and regional SO model for waters south of 50°S. These two models presented overall the lowest fitting and validation errors for the corresponding regions. This approach would not only provide more accurate retrievals of diatoms in the SO, but also overcome the overestimation of the global ABA_{Zpd} model in oligotrophic waters. However, applying two models generated a non-negligible offset between the SO and adjacent oceans (result not shown).

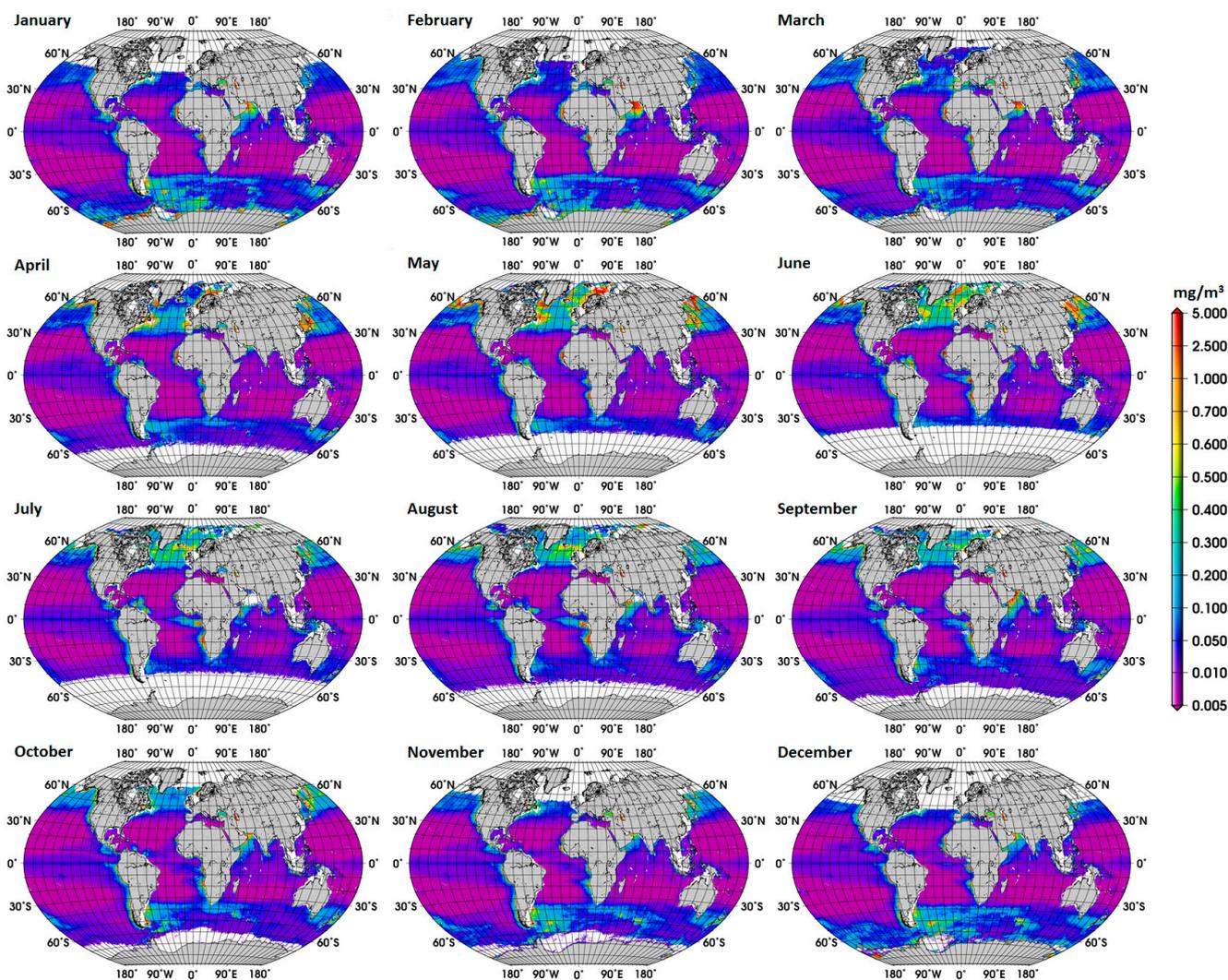
3.2. Satellite Retrieval of Diatoms Using ABA_{Zpd}

Acknowledging the uncertainties of the satellite Chla product, we first assessed the difference between the satellite retrievals of diatom abundance using the ABA_{Zpd} and the original ABA, for the SO and global oceans. As expected from the previous findings (Figure 2a), we observed that, on average, higher abundances of diatoms were retrieved with the ABA_{Zpd} than with the original ABA for the entire 2003–2013 period. For the SO, the concentration of diatoms calculated using the global ABA_{Zpd} is 0.074 mg/m³ and for the global oceans 0.070 mg/m³. In contrast, estimates of diatoms with the original ABA are 0.049 and 0.050 mg/m³, respectively. For comparison, the concentration of diatoms using the regional SO model is larger, 0.117 mg/m³. This evidence of the enhanced abundance of diatoms retrieved from the global ABA_{Zpd} model and from the regional SO model suggests that the production and export of carbon to the deep ocean might be larger than previously expected in the SO.

The new global climatology of diatom abundance is presented in Figure 4. The climatology for the SO is presented in the Supplementary Information (Figure S3). The general distribution of the global diatom abundance is in line with current knowledge on the distribution of diatoms, *i.e.*, higher concentrations of diatoms in the upwelling and coastal regions. Low concentrations of diatoms are observed in oligotrophic waters of the subtropical gyres and in high-nutrient low-chlorophyll (HNLC) waters, such as regions in the SO where waters are rich in macronutrients but are lacking in iron. There is also a clear seasonal cycle in the polar regions, with diatoms reaching the highest concentrations during their respective summer months, which is also observed in the climatology for the SO. Among other important patterns is the increase in diatom concentration from December to February and in September in the Arabian Sea. These observed patterns are associated with the Northeast (NE) and Southwest (SW) monsoons, respectively. According to Garrison *et al.* [33], the monsoon seasons are generally characterized by increased concentrations of diatoms, thus our result shows a consistency with the previous *in situ* study too.

The climatology mostly covers the spatial variability, within a limited temporal range, whereas the trend gives information for a longer period, and both are important information for the understanding of ocean biogeochemistry. The spatial variability of the linear trends of diatom abundance in the SO is high, and no significant trend was observed for most of the sub regions of the SO (results not shown). The overall the trend for the SO was 0.036 (dimensionless) (standard deviation = 12.84, p-value = 0.019). Clearly, a more detailed analysis is needed to investigate the main driving forces behind these trends.

Figure 4. Climatology of TChla_{Z_{pd}} of diatoms (mg/m³) for the months of January to December based on the period 2003–2013 retrieved using the ABA_{Z_{pd}} model. White areas correspond to waters with depths shallower than 200 m or without satellite information.



4. Conclusions

In conclusion, we have shown that the original ABA underestimates the diatom abundance in the Southern Ocean (SO). Our investigation revealed that diatoms in the SO might be more abundant than previous thought, possibly because (1) the lack of *in situ* phytoplankton pigment data, and that (2) the relationship between TChla and the *f*-Diatom in the SO is distinct from the global relationship.

We have developed a new global and regional ABA_{Z_{pd}} that significantly improve the representativeness of the retrievals in the SO. The mean absolute error (MAE) declined from 0.776 to 0.559 using the global ABA_{Z_{pd}}, improving by 28% the estimation of diatoms in the SO. The regional model further improved the MAE by 17% (MAE = 0.465) compared with the global ABA_{Z_{pd}} model. This was achieved by re-evaluating the ABA using a large data set of global phytoplankton pigment profiles spanning 24 years (1988–2012). Additionally, the ABA was further improved by considering the information in the Z_{pd}.

We have shown that the ideal global retrieval of diatoms combines the ABA_{Zpd} model fitted to the data set (excluding SO data, MAE = 0.883) with the regional SO model. However, applying two models generates an offset between the oceans, thus selective use of the global and the SO algorithms may be necessary depending on the objective of the application.

Satellite retrievals of PFTs are a useful tool for identifying and quantifying their presence in the oceans and in this study we have advanced our knowledge on the retrieval of diatoms from space by identifying limitations and developing improvements. Future studies should focus on optimizing the ABA method also for other PFTs.

Acknowledgments

We would like to thank all principal investigators and contributors for collecting the *in situ* data available in SeaBASS, MAREDAT, Lter database, BonusGoodHope project (LOV, Josephine Ras, Amelie Tale, Herve Claustre), KEOPS cruise (Herve Claustre) and from the other individual cruises mentioned in this paper. Data from the Palmer LTER data repository were supported by Office of Polar Programs, NSF Grants OPP-9011927, OPP-9632763 and OPP-0217282. We are also thankful to Werner Wosnok for the statistical advice and Marc Taylor for discussions. We acknowledge the Ocean Biology Processing Group of NASA and the NASA Goddard Space Flight Center's Ocean Data Processing System (ODPS) for the production and distribution of the MODIS data. We would like to thank the three anonymous reviewers, whose comments lead to a significantly improved manuscript. The present study was funded by the Helmholtz Association via the project Helmholtz-University Young Investigators Group PHYTOOPTICS in cooperation with the Institute of Environmental Physics (University Bremen), by the Total Foundation Group via project Phytoscope and by the Japan Aerospace Exploration Agency (JAXA) via the project Global Observation Mission-Climate. The first author is funded by the CAPES, Brazil, by the research grant BEX 3483/09-6.

Author Contributions

Mariana Soppa led the study, performed data analysis and wrote the draft version of the manuscript. Takafumi Hirata and Astrid Bracher supervised the work. Brenner Silva and Tilman Dinter contributed in statistical support and programming. Ilka Peeken provided the pigment data sets collected during ANT XVIII/2, ANT XXVI/3, ANT XXI/3, Meteor 55, Meteor 60 and Sonja Wiegmann was involved with laboratory analysis and sampling of ANT XXVIII/3, Sonne SO218, Merian 18-3 phytoplankton pigment data. All co-authors assisted in writing the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References and Notes

1. Hirata, T.; Hardman-Mountford, N. J.; Brewin, R.J.W.; Aiken, J.; Barlow, R.; Suzuki, K.; Isada, T.; Howell, E.; Hashioka, T.; Noguchi-Aita, M.; *et al.* Synoptic relationships between surface Chlorophyll-a and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences* **2011**, *8*, 311–327.

2. Rousseaux, C.S.; Gregg, W.W. Interannual variation in phytoplankton primary production at a global scale. *Remote Sens.* **2014**, *6*, 1–19.
3. Smetacek, V.; Klass, C.; Strass, V.H.; Assmy, P.; Montresor, M.; Cisewski, B.; Savoye, N.; Webb, A.; d'Ovidio F.; *et al.* Deep carbon export from a Southern Ocean iron-fertilized diatom bloom. *Nature* **2012**, *487*, 313–319.
4. Armbrust, E.V. The life of diatoms in the world's oceans. *Nature* **2009**, *459*, 185–192.
5. Leblanc, K.; Aristegui, J.; Armand, L.; Assmy, P.; Beker, B.; Bode, A.; Breton, E.; Cornet, V.; Gibson, J.; Gosselin, M.-P.; *et al.* A global diatom database—Abundance, biovolume and biomass in the world ocean. *Earth Syst. Sci. Data* **2012**, *4*, 149–165.
6. Wright, S.W.; van den Enden, R.L. Phytoplankton community structure and stocks in the East Antarctic marginal ice zone (BROKE survey, January–March 1996) determined by CHEMTAX analysis of HPLC pigment signatures. *Deep-Sea Res. Pt. II-Top. St. Oce.* **2000**, *47*, 2363–2400.
7. Arrigo, K.R.; Robinson, D.H.; Worthen, D.L.; Dunbar, R.B.; DiTullio, G.R.; VanWoert, M.; Lizotte, M.P. Phytoplankton community structure and the drawdown of nutrients and CO₂ in the Southern Ocean. *Science* **1999**, *283*, 365–367.
8. Bracher, A.U.; Kroon, B.M.A.; Lucas, M.I. Primary production, physiological state and composition of phytoplankton in the Atlantic Sector of the Southern Ocean. *Mar. Ecol. Prog. Ser.* **1999**, *190*, 1–16.
9. Alvain, S.; Moulin, C.; Dandonneau, Y.; Bréon, F.M. Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. *Deep-Sea Res. Pt. I-Oceanog. Res.* **2005**, *52*, 1989–2004.
10. Bracher, A.; Vountas, M.; Dinter, T.; Burrows, J.P.; Röttgers, R.; Peeken, I. Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences* **2009**, *6*, 751–764.
11. Devred, E.; Sathyendranath, S.; Stuart, V.; Maass, H.; Ulloa, O.; Platt, T. A two-component model of phytoplankton absorption in the open ocean: Theory and applications. *J. Geophys. Res.-Oceans* **2006**, *111*, doi:10.1029/2005JC002880.
12. Hirata, T.; Aiken, J.; Hardman-Mountford, N.; Smyth, T.J.; Barlow, R.G. An absorption model to determine phytoplankton size classes from satellite ocean colour. *Remote Sens. Environ.* **2008**, *112*, 3153–3159.
13. Brewin, R.J.W.; Sathyendranath, S.; Hirata, T.; Lavender, S.J.; Barciela, R.M.; Hardman-Mountford, N.J. A three-component model of phytoplankton size class for the Atlantic Ocean. *Ecol. Model.* **2010**, *221*, 1472–1483.
14. Gordon, H.R.; McCluney, W.R. Estimation of depth of sunlight penetration in sea for remote sensing. *Appl. Opt.* **1975**, *14*, 413–416.
15. Uitz, J.; Claustre, H.; Morel, A.; Hooker, S.B. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *J. Geophys. Res.-Oceans* **2006**, *111*, doi:10.1029/2005JC003207.
16. Hyde, K.J.W.; O'Reilly, J.E.; Oviatt, C.A.; Validation of SeaWiFS chlorophyll a in Massachusetts Bay. *Cont. Shelf Res.* **2007**, *27*, 1677–1691.

17. Vidussi, F.; Claustre, H.; Manca, B.B.; Luchetta, A.; Marty, J.C. Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *J. Geophys. Res.-Oceans* **2001**, *106*, 19939–19956.
18. Werdell, P.J.; Bailey, S.; Fargion, G.; Pietras, C.; Knobelspiesse, K.; Feldman, G.; McClain, C. Unique data repository facilitates ocean color satellite validation. *EOS. Trans. AGU* **2003**, *84*, 377–387.
19. Peloquin, J.; Swan, C.; Gruber, N.; Vogt, M.; Claustre, H.; Ras, J.; Uitz, J.; Barlow, R.; Behrenfeld, M.; Bidigare, R. The MAREDAT global database of high performance liquid chromatography marine pigment measurements. *Earth Syst. Sci. Data* **2013**, *5*, 109–123.
20. Uitz, J.; Claustre, H.; Griffiths, F.B.; Ras, J.; Garcia, N.; Sandroni, V. A phytoplankton class-specific primary production model applied to the Kerguelen Islands region (Southern Ocean). *Deep-Sea Res. Pt. I-Oceanog. Res.* **2009**, *56*, 541–560.
21. Cheah, W.; Taylor, B.B.; Wiegmann, S.; Raimund, S.; Krahnemann, G.; Quack, B.; Bracher, A. Photophysiological state of natural phytoplankton communities in the South China Sea and Sulu Sea. *Biogeosciences Discuss.* **2013**, *10*, 12115–12153.
22. Hoffmann, L.; Peeken, I.; Lochte, K.; Assmy, P.; Veldhuis, M. Different reactions of Southern Ocean phytoplankton size classes to iron fertilization. *Limnol. Oceanogr.* **2006**, *51*, 1217–1229.
23. Taylor, B.B.; Torrecilla, E.; Bernhardt, A.; Taylor, M.H.; Peeken, I.; Röttgers, R.; Piera, J.; Bracher, A. Bio-optical provinces in the eastern Atlantic Ocean and their biogeographical relevance. *Biogeosciences Discuss.* **2011**, *8*, 3609–3629.
24. Amante, C. Eakins, B.W. *Etopo1 1 ARC-Minute Global Relief Model: Procedures, Data Sources and Analysis*; U S Government: Washington, DC, USA, 2009; p. 19.
25. OceanColor Web. Available online: <http://oceancolor.gsfc.nasa.gov/> (accessed on 10 February 2014).
26. Gordon, H.R.; Clark, D.K. Remote sensing optical-properties of a stratified ocean: An improved interpretation. *Appl. Opt.* **1980**, *19*, 3428–3430.
27. Morel, A.; Maritorena, S. Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.-Oceans* **2001**, *106*, 7163–7180.
28. Lee, Z.; Du, K.; Arnone, R.; Liew, S.; Penta, B. Penetration of solar radiation in the upper ocean: A numerical model for oceanic and coastal waters. *J. Geophys. Res.-Oceans* **2005**, *110*, doi:10.1029/2004JC002780.
29. Willmott, C.J.; Matsuura, K.; Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82.
30. Brewin, R.J.; Sathyendranath, S.; Lange, P.K.; Tilstone, G. Comparison of two methods to derive the size-structure of natural populations of phytoplankton. *Deep-Sea Res. Pt. I-Oceanog. Res.* **2014**, *85*, 72–79.
31. Fujiwara, A.; Hirawake, T.; Suzuki, K.; Imai, I.; Saitoh, S.I. Timing of sea ice retreat can alter phytoplankton community structure in the western Arctic Ocean. *Biogeosciences* **2014**, *11*, 1705–1716.
32. Sadeghi, A.; Dinter, T.; Vountas, M.; Taylor, B.; Peeken, I.; Altenburg Soppa, M.; Bracher, A. Improvement to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data. *Ocean Sci.* **2012**, *8*, 1055–1070.

33. Garrison, D.L.; Gowing, M.M.; Hughes, M.P.; Campbell, L.; Caron, D.A.; Dennett, M.R.; Shalapyonok, A.; Olson, R.J.; Landry, M.R.; Brown, S.L.; *et al.* Microbial food web structure in the Arabian Sea: A US JGOFS study. *Deep-Sea Res. Pt. II-Top. St. Oce.* **2000**, *47*, 1387–1422.

© 2014 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 license (<http://creativecommons.org/licenses/by/4.0/>).