

## RESEARCH ARTICLE

10.1029/2018JC014329

## Key Points:

- Multivariate assimilation of satellite chlorophyll data improves simulation of phytoplankton functional groups and influences them differently
- Small phytoplankton is weakly deteriorated in the Southern Ocean, while diatoms are improved globally
- Regional variability of assimilation leads to stronger improvement at midlatitudes and equator than at high latitudes

## Correspondence to:

H. K. Pradhan,  
hpradhan@awi.de

## Citation:

Pradhan, H. K., Völker, C., Losa, S. N., Bracher, A., & Nerger, L. (2019). Assimilation of global total chlorophyll OC-CCI data and its impact on individual phytoplankton fields. *Journal of Geophysical Research: Oceans*, 124, 470–490. <https://doi.org/10.1029/2018JC014329>

Received 2 JUL 2018

Accepted 19 DEC 2018

Accepted article online 21 DEC 2018

Published online 21 JAN 2019

## Assimilation of Global Total Chlorophyll OC-CCI Data and Its Impact on Individual Phytoplankton Fields

H. K. Pradhan<sup>1</sup> , C. Völker<sup>1</sup> , S. N. Losa<sup>1,2</sup> , A. Bracher<sup>1,3</sup> , and L. Nerger<sup>1</sup> 

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

**Abstract** The coupled ocean circulation-ecosystem model MITgcm-REcoM2 is used to simulate biogeochemical variables in a global configuration. The ecosystem model REcoM2 simulates two phytoplankton groups, diatoms and small phytoplankton, using a quota formulation with variable carbon, nitrogen, and chlorophyll contents of the cells. To improve the simulation of the phytoplankton variables, chlorophyll-a data from the European Space Agency Ocean-Color Climate Change Initiative (OC-CCI) for 2008 and 2009 are assimilated with an ensemble Kalman filter. Utilizing the multivariate cross covariances estimated by the model ensemble, the assimilation constrains all model variables describing the two phytoplankton groups. Evaluating the assimilation results against the satellite data product SynSenPFT shows an improvement of total chlorophyll and more importantly of individual phytoplankton groups. The assimilation improves both phytoplankton groups in the tropical and midlatitude regions, whereas the assimilation has a mixed response in the high-latitude regions. Diatoms are most improved in the major ocean basins, whereas small phytoplankton show small deteriorations in the Southern Ocean. The improvement of diatoms is larger when the multivariate assimilation is computed using the ensemble-estimated cross covariances between total chlorophyll and the phytoplankton groups than when the groups are updated so that their ratio to total chlorophyll is preserved. The comparison with in situ observations shows that the correlation of the simulated chlorophyll of both phytoplankton groups with these data is increased whereas the bias and error are decreased. Overall, the multivariate assimilation of total chlorophyll modifies the two phytoplankton groups separately, even though the sum of their individual chlorophyll concentrations represents the total chlorophyll.

**Plain Language Summary** Different types of plankton are simulated globally with ocean ecosystem models. To further increase their prediction quality, we combine the model with satellite observations of chlorophyll using modern methods called data assimilation. This method allows us not only to improve the modeled total chlorophyll but also the simulation of the different plankton types. Further, we can fill gaps in the satellite data that results, for example, from clouds. Thus, we are able to better predict the ocean ecosystem, which in turn helps to understand climate change patterns and carbon cycle processes.

### 1. Introduction

The biogeochemistry of the ocean plays an important role in the carbon cycle and is expected to be affected by climate change (Friedlingstein et al., 2001). The available resources to study ocean biogeochemistry and its role in climate are satellite data, in situ observations, and numerical models, each with specific uncertainties and limitations. One of the largest sources of information to test our understanding of marine biogeochemistry is ocean chlorophyll estimates available from satellite data, which provide global coverage of the first optical depth of the ocean but also miss data mainly due to clouds, aerosols, sun glint, and below the first optical depth (e.g., Hammond et al., 2017). Errors in the satellite chlorophyll data due to the indirect measurement via remote sensing and exploration of the optical signal can be higher than 35% (e.g., Maritorena et al., 2010). However, data products are now available with uncertainties on a pixel-by-pixel basis (Jackson et al., 2017). In situ observations are more accurate but very sparse in time and space (e.g., see Valente et al., 2016). To derive inferences on the ocean carbon cycle from chlorophyll data, from either in situ measurements or satellites, one further needs an assumption on the carbon-to-chlorophyll ratio of phytoplankton, which can vary over an order of magnitude. Biogeochemical models can give full spatial and temporal coverage of the world ocean and usually include carbon state variables directly, but their

accuracy is limited by the model resolution, by inherent approximations and parameterizations, and by uncertainties in model forcing and initial conditions. Data assimilation combines the information from model dynamics and observations. It can fill observational data gaps and can transfer information from the observations to unobserved model variables.

Several recent studies have focused on the assimilation of satellite ocean chlorophyll to coupled ocean-biogeochemical models both on regional and global scales. Natvik and Evensen (2003) successfully assimilated remotely sensed chlorophyll data in the North Atlantic into a model. The assimilation of surface chlorophyll improved phosphate concentrations in the eastern Mediterranean Sea (Triantafyllou et al., 2007). Ourmières et al. (2009) assimilated nitrate climatology into a model of the North Atlantic with the objective to estimate basin-scale patterns of oceanic primary production. Ciavatta et al. (2011) assimilated Moderate Resolution Imaging Spectroradiometer (MODIS) chlorophyll data for the year 2006 in the western English Channel. Rousseaux and Gregg (2012) studied the effect of climate variability on the phytoplankton community in the Pacific Ocean assimilating surface chlorophyll from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS). Simon et al. (2015) assimilated both physical and biogeochemical data sets for estimating phytoplankton and zooplankton mortality rate in the North Atlantic and Arctic Oceans on a weekly basis. Recently, Ciavatta et al. (2016) generated a reanalysis for the northwest European Shelf by assimilating satellite total chlorophyll data. Further, Ciavatta et al. (2018), on the same domain, assimilated a data set of phytoplankton functional types (PFTs), which was generated to be compatible with their model. They found that the phytoplankton size class assimilation improved the phytoplankton community structure and total chlorophyll, compared to a simulation which only assimilated total chlorophyll data.

On a global scale, Tjiputra et al. (2007) assimilated seasonal chlorophyll and nutrient data giving a three-dimensional adjoint model to analyze the sensitivity of model-generated chlorophyll and optimize ecosystem parameters. Nerger and Gregg (2007) demonstrated that the assimilation of SeaWiFS chlorophyll data with an ensemble Kalman filter (EnKF) improved the estimates of primary production and generated a 7-year reanalysis of surface chlorophyll concentrations in a global functional type model, while Nerger and Gregg (2008) further corrected model biases. Gregg (2008) assimilated daily surface chlorophyll from SeaWiFS and MODIS-Aqua into a global model using a conditional relaxation method. The study discusses the assimilation impact in different oceanic basins where both bias and uncertainty were reduced by the assimilation. Ford et al. (2012) assimilated chlorophyll-*a* improving the surface chlorophyll and updating other biogeochemical variables through a mechanistic nitrogen balancing scheme. Gregg and Rousseaux (2014) discussed the global decadal trends of pelagic ocean chlorophyll, and Rousseaux and Gregg (2015) assessed the trends in phytoplankton composition. Both studies assimilated chlorophyll, but they did not discuss the effect of the assimilation itself. Xiao and Friedrichs (2014) assimilated actual and synthetic satellite data for particulate organic carbon, size-differentiated chlorophyll, and total chlorophyll into a one-dimensional model for optimizing biogeochemical parameters. Ford and Barciela (2017) generated a reanalysis by assimilating two different satellite chlorophyll data sets from the European Space Agency Ocean-Color Climate Change Initiative (OC-CCI, <http://www.esa-oceancolour-cci.org/>) and GlobColour (Maritorena et al., 2010; <http://www.globcolour.info/>). They found that both data sets have a similar influence in the assimilation, but OC-CCI data had a better spatial coverage than GlobColour data. Basically, the above studies demonstrate successful assimilation on global and regional scales improving at least the observed fields. Recently, Skákala et al. (2018) assimilated PFTs, univariately on the northwest European Shelf improving the PFTs and total chlorophyll and discussed the impact on unassimilated variables. When total chlorophyll data are assimilated into a model simulating multiple phytoplankton groups, the correction of the total chlorophyll needs to be distributed over the different groups. There are two approaches to achieve this. The first one was applied, for example, by Gregg (2008), who used the condition to keep the ratio of the chlorophyll from each phytoplankton group to total chlorophyll constant. So the relative abundance of the phytoplankton groups is not changed by the assimilation but only by the model dynamics. The same condition was used by Nerger and Gregg (2007) with an EnKF. This approach is useful when one does not have additional information about the phytoplankton community. However, a better assimilation result should be expected from the application of multivariate ensemble assimilation that allows the relative abundances to change. In this case one uses cross covariances between total chlorophyll and the PFTs, which are dynamically estimated by the ensemble of model states. Thus, an increment of total chlorophyll will be transferred to increments of unobserved PFTs, but possibly also to chlorophyll

concentrations below the ocean surface. Apart from correcting different PFTs of a model, the multivariate assimilation can also be used to update other model variables like nutrients.

Only a few assimilation studies have considered the multivariate effect on phytoplankton groups since most studies only use models that simulate a single phytoplankton group (e.g., Doron et al., 2013; Ford et al., 2012; Ford & Barciela, 2017; Hu et al., 2012). In a first study applying multivariate ensemble assimilation examining phytoplankton groups, Ciavatta et al. (2011) assimilated chlorophyll data in the western English Channel for 2006. Their assimilation improved most of the biogeochemical variables, both assimilated and nonassimilated, but the model estimate of dinoflagellates was not improved. The improvements happened at the assimilation times, while there were no significant improvements for weekly forecasts computed in between the assimilation times. A reverse strategy to Ciavatta et al. (2011) was taken by Ciavatta et al. (2018). They assimilated a data set of PFTs, which was generated to be compatible with their model, and assessed the assimilation effect on both the PFTs and total chlorophyll. They found that the PFTs assimilation improved the phytoplankton community structure and total chlorophyll, compared to a simulation which assimilated total chlorophyll data. However, diatoms and nanophytoplankton were overestimated. Shulman et al. (2013) performed multivariate assimilation of the underwater light field, that is, optical properties, which improved PFTs and nitrate field in the Monterey Bay region.

To gain more insight into the multivariate assimilation effects on PFTs, this study assesses the assimilation of satellite chlorophyll-a (hereafter chlorophyll) data in a global configuration of the Regulated Ecosystem Model-2 (REcoM2; see Hauck et al., 2013) coupled with the MITgcm ocean circulation model (MITgcm Group, 2018). The assimilation is conducted every 5 days over the years 2008 and 2009 with an EnKF which is configured so that the assimilation directly changes all variables that describe the two PFTs (diatoms and small phytoplankton) of REcoM2 through cross covariances. In addition, a configuration of the EnKF is used in which the PFT concentrations are changed so that their ratio to the total chlorophyll is preserved. The particular aim of the study is to assess how the multivariate assimilation influences the chlorophyll concentration of the two distinct PFTs of REcoM2 and how this effect varies regionally.

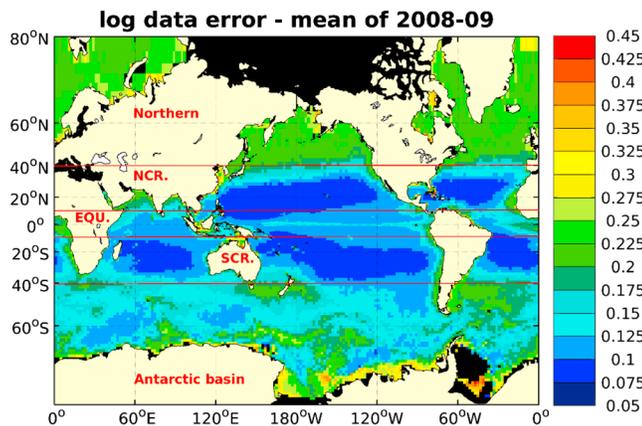
The paper is organized as follows: The model, data assimilation technique, data utilized, and the setup of the numerical experiments are described in section 2. The results of the assimilation are described in section 3, while section 4 discusses the success of the method and how the model state is affected. Section 5 provides the general conclusions.

## 2. Methods

### 2.1. The MITgcm-REcoM2 Model

The physical model utilized for our study is the MITgcm (Marshall et al., 1997; MITgcm Group, 2018). We use a global model configuration excluding the Arctic basin, extending from 80°N to 79°S. The resolution along longitude is a constant 2°. In the latitudinal direction the resolution is 2° in the Northern Hemisphere and is getting gradually finer to 0.38° from the equator toward the Antarctic, with additional increased resolution in a band around the equator. In the vertical it has 30 layers varying in thickness from 10 m at the surface to 500 m toward the bottom. The model configuration includes a prognostic sea ice component. At 80°N the model has a solid boundary; that is, there is a no-flux boundary there.

The ecosystem model REcoM2 belongs to a class of quota models in which the photo-acclimation is based on Geider et al. (1998). Early versions of REcoM2 were developed by Schartau et al. (2007) and Hohn (2009). It has compartments for phytoplankton, zooplankton, detritus, and main inorganic and dissolved nutrients. The highest trophic level in the model is zooplankton. Overall, it has 21 biogeochemical tracers. REcoM2 is coupled with MITgcm (Losch et al., 2008) and is used for large-scale simulations with focus on the Southern Ocean (Hauck et al., 2016; Hauck & Völker, 2015). REcoM2 simulates two phytoplankton groups: small phytoplankton and diatoms. The two groups are parametrized in such a way that the diatom group has a higher maximum growth rate and a larger initial slope of photosynthesis-irradiance curve than the small phytoplankton. Although the nutrient uptake half-saturation constant is higher for diatoms, this gives them a competitive advantage under high nutrient conditions. The small phytoplankton group is assumed to contain a small (in the current run 10%) fixed fraction of calcifiers (coccolithophores), with a constant calcium carbonate/organic carbon ratio. The minimum concentrations of the biogeochemical variables in the model



**Figure 1.** Per pixel logarithmic observation error of total chlorophyll provided by the Ocean-Color Climate Change Initiative data set averaged in the model grid and for 2008 and 2009. The global domain is divided into five regions marked by the red lines: Northern, northern central region (NCR), equatorial (EQU.), southern central region (SCR), and Antarctic basin.

are limited to a value of  $10^{-4}$  in the respective model units, which are millimoles per cubic meter for carbon- and nitrogen-based biomass and milligrams per cubic meter for chlorophyll. Detailed descriptions about the coupling and configuration can be found in Hauck et al. (2013).

## 2.2. Data Assimilation

For the data assimilation, the Parallel Data Assimilation Framework (Nerger & Hiller, 2013; <http://pdaf.awi.de>) was coupled to MITgcm-REcoM2. The assimilation framework provides the environment for ensemble simulations and different ensemble-based Kalman filters. The assimilation methodology used here is the local error-subspace transform Kalman filter (LESTKF; Nerger et al., 2012), which was, for example, applied by Chen et al. (2017) for assimilating sea ice concentration and thickness. Compared to the classical EnKF (Evensen, 1994), the LESTKF is computationally more efficient because it directly accounts for the fact that the degrees of freedom for the assimilation are given by the ensemble size. Further, the LESTKF has lower sampling error than the EnKF, as it does not require to perturb the observations (see Nerger et al., 2007, for a discussion on this aspect for the related Singular

‘Evolutive’ Interpolated Kalman (SEIK) filter). The analysis step is performed without actually restarting the model program in our data assimilation setup.

The localization scheme updates each vertical column of the model separately using only observations within a prescribed horizontal influence radius of  $5^\circ$  around the water column. Within this radius the observation influence is damped toward 0 with increasing distance. Next to the horizontal localization, a vertical localization is introduced. For this, the assimilation increment is first computed for the full vertical column. Then, the increment is tapered with a linear weight function of depth so that it is reduced to 0 at a depth of 75 m.

Given that chlorophyll is lognormally distributed (Campbell, 1995) the data assimilation is performed using log-transformed concentrations. For the ensemble states this is achieved by the direct computation of the logarithm. The transformation assures statistical consistency and positiveness of the concentrations (Nerger & Gregg, 2007). Since the assimilation of logarithmic concentrations can result in larger increments compared to an assimilation of actual concentrations, the size of the logarithmic assimilation increments is limited to  $<1.0$ . Given that the assimilation uses the natural logarithm this implies that actual concentrations are changed at most by a factor of 2.718.

## 2.3. Assimilation and Evaluation Data

### 2.3.1. Assimilated Satellite Data of Total Chlorophyll Concentration

The assimilated chlorophyll data were taken from the release version 3.1 of the OC-CCI (Sathyendranath et al., 2018; <https://doi.org/10.5285/9c334f6e6d424a708cf3c4cf0c6a53f5>) of the European Space Agency (<http://www.esa-oceancolour-cci.org/>). The data product comprises globally merged Medium-Resolution Imaging Spectrometer, Aqua-MODIS, SeaWiFS, and Visible Infrared Imaging Radiometer Suite data. We used the 5-day composite of chlorophyll concentrations, error information in the form of the logarithmic root-mean-square (RMS) deviation and the bias. Figure 1 shows the spatially varying RMS deviation averaged for 2008 and 2009, averaged on the model grid. The error is higher near to coastal regions, continental shelves, and higher latitudes, whereas regions away from the shelf or deeper oceans have the smallest error. The 5-day composite error information is used to specify the observation error for our assimilation experiments. To obtain the logarithmic concentration and associated errors for the 5-day composites used in the assimilation, we followed the averaging procedure described in the OC-CCI product manual (Grant et al., 2015; see also Ciavatta et al., 2016).

### 2.3.2. Satellite-Derived Chlorophyll Concentration for Specific PFTs

For the global evaluation of the two phytoplankton groups in our model, the Synergistic hyper- and multi-spectral satellite PFT (SynSenPFT) product by Losa et al. (2017a) was used. This is a synergistic product of two satellite retrieval algorithms. The large-scale information ( $0.5^\circ$  latitude times  $0.5^\circ$  longitude and weekly resolution) on the three phytoplankton-types diatoms, coccolithophores, and cyanobacteria is derived with

the analytical method PhytoDOAS (Bracher et al., 2009; Bracher, Dinter, et al., 2017; Sadeghi et al., 2012), which resolves the spectral imprints of each individual group from the spectrally highly resolved measurements of the sensor SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCIAMACHY) mounted on Envisat. This is combined with small-scale information (4 km by 4 km and daily resolution) on the three groups obtained using the empirical abundance-based OC-PFT (Ocean Color-Phytoplankton Functional Type) method (Hirata et al., 2011; Soppa et al., 2014, 2016; supplement in Losa et al., 2017a). The OC-PFT method uses a statistical model relating a fraction of retrieved PFTs to total chlorophyll. For the SynSenPFT input data the OC-PFT method is applied to total chlorophyll from the OC-CCI. Both data types are then combined into the synergistic SynSenPFT product using optimal interpolation.

Because the OC-PFT estimates depend on OC-CCI total chlorophyll concentrations, which are assimilated here, the SynSenPFT product (Losa et al., 2017a; products available at Losa et al., 2017b) is not a completely independent data set. However, the partitioning of total chlorophyll among different PFTs in the OC-PFT algorithm is fully independent from the model dynamics in the MITgcm-REcoM. Furthermore, SynSenPFT is a combined product of OC-PFT and PhytoDOAS PFT chlorophyll. The latter uses a PFT-specific spectral optical “fingerprint” for distinguishing between various phytoplankton types which is completely different from the empirical OC-PFT algorithm. Accordingly, the combined SynSenPFT data set depends only partially on the OC-CCI total chlorophyll data product.

The SynSenPFT data product provides daily chlorophyll concentrations for diatoms, coccolithophores, and cyanobacteria for the global ocean on a 4-km sinusoidal grid for the period from August 2002 to March 2012. To compare the data assimilation results with the SynSenPFT data, 5-day composites were computed and then averaged on the model grid for 2008 and 2009. In our evaluation, the small phytoplankton of REcoM2 is compared with the sum of coccolithophores and cyanobacteria of the SynSenPFT data, which is necessarily an approximation since also other small-sized groups can contribute significantly to this size class (e.g., other groups of prymnesiophytes, chlorophytes, and pelagophytes). In our evaluation we have excluded the period from 15 December 2008 to 15 January 2009 because the number of hyperspectral input data for the PhytoDOAS algorithm was very low after removing contaminants (due to heating of SCIAMACHY detectors during this period; for more details see [https://earth.esa.int/documents/700255/708683/RMF\\_0140\\_SCI\\_NL\\_\\_1P\\_v1.1\\_Dec2016.pdf](https://earth.esa.int/documents/700255/708683/RMF_0140_SCI_NL__1P_v1.1_Dec2016.pdf)) so that the SynSenPFT data within this time frame was mostly determined by the OC-PFT chlorophyll concentrations.

The uncertainties of the SynSenPFT data have implications on their applicability to evaluate the assimilation output. Overall, uncertainties result from the empirical OC-PFT algorithm, which uses the already uncertain OC-CCI total chlorophyll concentrations and the error approximations for the combination of both OC-PFT and PhytoDOAS data. In particular, the PhytoDOAS errors are assumed to be constant over space and time, while they are expected to be spatially and temporally varying. Further, there are representation errors due to different temporal and spatial scales represented in the two input data sets for SynSenPFT. Another representation error results from the different grid resolutions used in our model compared to the 4-km resolution of SynSenPFT. In addition, there is a mismatch between the three PFTs represented in the SynSenPFT data compared to the two PFTs simulated by REcoM2. A detailed discussion of the possible data errors in SynSenPFT is provided by Losa et al. (2017a). An indication of the data quality is provided by the comparison to in situ data using scatterplots discussed in section 3.3.

### 2.3.3. In Situ Observations of Total and PFT-Specific Chlorophyll Concentrations

Further evaluation of the model results was performed with a large in situ data set by Soppa et al. (2017). The data set contains chlorophyll concentrations for diatoms, haptophytes, and prokaryotes. It has been derived by using the Diagnostic Pigment Analysis by Vidussi et al. (2001) and Uitz et al. (2006) modified as in Hirata et al. (2011) and Brewin et al. (2015) of in situ phytoplankton pigment data based on high-precision liquid chromatography and compiled from several databases and individual cruises (see Losa et al., 2017a). In our analysis, we approximate the in situ small phytoplankton with the sum of haptophytes and prokaryotes. The in situ data set contains also total chlorophyll concentrations that are derived from the sum of monovinyl chlorophyll-a, chlorophyllide-a, and divinyl-chlorophyll-a concentrations. Only measurements with concentrations of at least  $0.01 \text{ mg/m}^3$  were taken into account for our analysis since lower values appear to be unrealistic (see Losa et al., 2017a). For the comparison with the model fields, the nearest model grid point was used. Further, the data over a 5-day window is mapped to the end of the window when the assimilation is performed. For the assimilation period extending through the years 2008 and

**Table 1**  
*RMS Errors for Model Small Phytoplankton Compared to SynSenPFT Data Separated Over Zonal Regions for the Free Run, Assimilation Forecast, and Analysis*

Domain	Small phytoplankton				
	Free run	Cross covariance		Ratio preserving	
		Forecast	Analysis	Forecast	Analysis
Global	0.609	0.539	0.507	0.531	<b>0.503</b>
Northern	0.948	0.927	0.901	0.918	<b>0.884</b>
NCR	0.47	0.424	<b>0.390</b>	0.428	0.413
Equatorial	0.427	0.345	<b>0.281</b>	0.328	0.285
SCR	0.417	0.349	<b>0.318</b>	0.348	0.328
Antarctic basin	0.77	0.679	0.649	0.643	<b>0.598</b>

*Note.* RMS errors are shown for the multivariate updates with cross covariance and preserving the ratio of the PFTs. The bold values indicate the smallest value in each region. RMS = root-mean-square; NCR = northern central region; SCR = southern central region.

2009, the in situ data provide 1,077 points of total chlorophyll concentrations to be compared to the model simulations.

The in situ data are not completely independent from the SynSenPFT data. The fraction which is not colocated with OC-CCI data was used to determine the empirical functions used in the OC-PFT method. However, the colocated fraction of the data was only used in the validation of the SynSenPFT data set and is hence independent. Also, for the in situ data there are representation errors due to differences in temporal and spatial scales which limit the comparison of these data with the model results.

#### 2.4. Simulation Strategy

To prepare the assimilation, the years 2003 to 2006 were used to spin-up the coupled model (MITgcm-REcoM2) with a single ensemble member. The initial conditions at the start of the simulation (from January 2003) are from World Ocean Atlas 2009 (Levitus et al., 2010) temperature, salinity and, macronutrients (Dissolve Inorganic Nitrogen/Dissolve

Inorganic Nitrogen/Si) (Garcia et al., 2010), Global Data Analysis Project values for Dissolve Inorganic Carbon, and alkalinity (Key et al., 2004). Iron was initialized with concentrations obtained from a previous run by the PISCES model (Aumont et al., 2003). All other tracers were initialized with arbitrary small values. Atmospheric forcing are from Coordinated Ocean-Ice Reference Experiment (CORE) forcing (Large & Yeager, 2004), and a dust deposition field to calculate iron flux is from Mahowald (2003). For the year 2007 we perform an ensemble spin-up. The data assimilation was applied for 2 years from January 2008 to December 2009.

To generate ensemble members, we perturbed sensitive biogeochemistry parameters of the ecosystem model. Following earlier studies (Ciavatta et al., 2016; Doron et al., 2013; Hu et al., 2012; Jones et al., 2016), where the authors tried to account for model deficiencies due to uncertainties in the biogeochemical model parameter specification, we introduced some stochasticity to the model by perturbing eight biogeochemical parameters. In particular, we perturbed the chlorophyll degradation rate, the initial slope of the photosynthesis-irradiation curve, and the maximum specific rate of photosynthesis for both phytoplankton groups. In addition, we perturbed the maximum grazing rate and the grazing efficiency of the zooplankton as the chlorophyll concentrations also showed sensitivity to these parameters. The parameters were jittered assuming a lognormal distribution with a relative variance of 0.125 for all parameters. Using the perturbed parameters, an ensemble of 20 states was generated by a spin-up run over the year 2007 referred to as ensemble spin-up.

The data assimilation experiment was conducted with 5-day forecast/analysis cycles with the previously mentioned horizontal and vertical localization. To stabilize the data assimilation, the model error covariances are inflated in each analysis step using a forgetting factor of 0.8; that is, the inflation factor is the inverse of this value.

To maintain the stoichiometry, the multivariate data assimilation directly updates the eight fields of the biogeochemical model that describe the two phytoplankton groups. For both small phytoplankton and diatoms these are the phytoplankton content of carbon, nitrogen, and chlorophyll. Further calcium carbonate is updated for small phytoplankton and biogenic silica for diatoms. Other variables of REcoM2, like nutrients, zooplankton, and detritus are not included in the state vector and are hence only influenced indirectly by the assimilation via the model dynamics. This configuration of the state vector was a design decision which allows us to focus on the assimilation effect on the phytoplankton. Two configurations of the filter update were used: in the first, the eight variables describing the PFTs were updated through the ensemble-estimated cross covariances to total chlorophyll. In the second configuration the assimilation increment to total chlorophyll was distributed over the two phytoplankton groups so that their ratio to the total chlorophyll was preserved. The discussion below focuses on the first approach, which yielded better estimates, while the assessment of ratio-preserving approach is only discussed for the SynSenPFT data in Tables 1 and 2.

**Table 2**  
RMS Errors for Model Diatoms Compared to SynSenPFT Data Separated Over Zonal Regions for the Free Run, Assimilation Forecast, and Analysis Analogous to Table 1

Domain	Diatoms				
	Free run	Cross covariance		Ratio preserving	
		Forecast	Analysis	Forecast	Analysis
Global	1.96	1.75	<b>1.71</b>	1.94	1.91
Northern	1.77	1.52	<b>1.44</b>	1.71	1.65
NCR	2.11	1.89	<b>1.88</b>	2.08	2.07
Equatorial	1.91	1.70	<b>1.65</b>	1.89	1.85
SCR	2.2	2.01	<b>1.99</b>	2.19	2.18
Antarctic basin	1.66	1.44	<b>1.42</b>	1.63	1.6

Note. RMS = root-mean-square; NCR = northern central region; SCR = southern central region.

### 3. Results

To discuss the results of the experiments, we first focus on the total chlorophyll. Subsequently we discuss the effect of the assimilation on the two phytoplankton groups and validate the assimilation performance with satellite and in situ data.

#### 3.1. Assimilation Effect on Total Chlorophyll

As an example of the influence of the data assimilation on total chlorophyll, Figure 2 shows the concentration in the topmost model layer in milligrams per cubic meter from the (a) free run and the (b) assimilation experiment on 20 April 2018 during the spring bloom in the Northern Hemisphere. Further, the 5-day composites for the assimilated OC-CCI total chlorophyll data and the sum of PFTs chlorophyll from SynSenPFT data projected onto the model grid are shown in Figures 2a and 2b. For the assimilation experiment the analysis state, that is, the concentration directly after computing the analysis, is shown. Figures 2e and 2f show

the difference between free run and OC-CCI data and between assimilation run and OC-CCI data. The free run reproduces the main spatial patterns of the OC-CCI data. A particular feature of the model is visible in the Southern Hemisphere where the model shows concentrations of up to  $0.225 \text{ mg/m}^3$  around  $40^\circ\text{S}$  and increased concentrations of around  $0.5 \text{ mg/m}^3$  as far south as  $65^\circ\text{S}$ . In contrast, the satellite data show a band of higher concentrations of up to about  $0.5 \text{ mg/m}^3$  around  $40^\circ\text{S}$  and lower concentrations further south. Due to the light conditions, the satellite data do not cover the region south of  $60^\circ\text{S}$  at this time of the year.

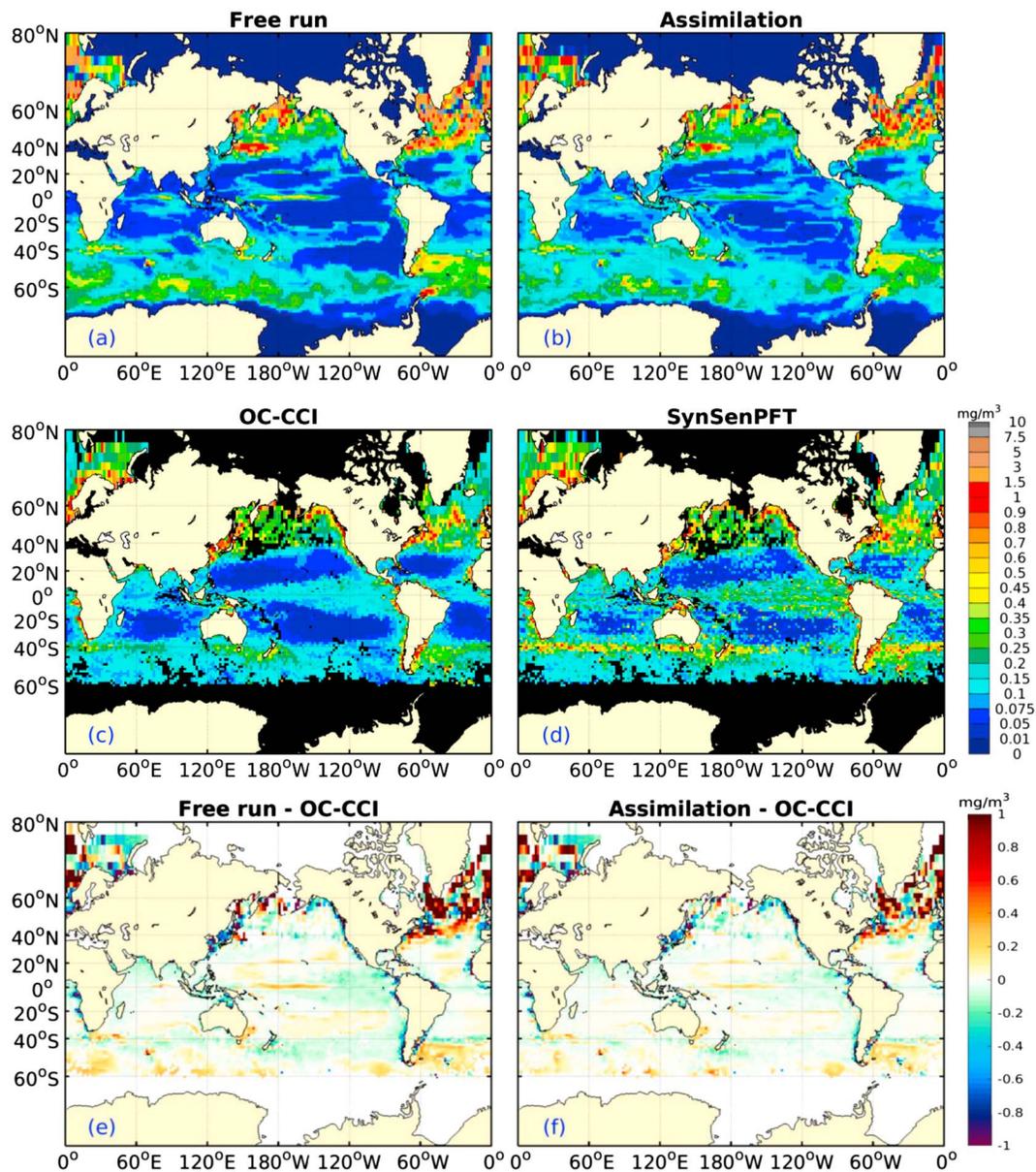
The chlorophyll concentration field from the assimilation experiment is closer to the satellite data. In the northern Atlantic and Pacific, in the equatorial region and the Southern Ocean the assimilation reduces the concentrations, bringing the model closer to the satellite data. The assimilation also improves regions with lower concentrations by increasing the chlorophyll concentrations. These are the Yellow Sea, the Gulf of Oman, and off the coast of Peru. The low-concentration regions of the subpolar gyres north and south of the equator in the Atlantic and Pacific and in the Indian Ocean between the equator and  $40^\circ\text{S}$  are neither improved nor deteriorated.

Figure 2d also shows the sum of the three functional types of the SynSenPFT data as an approximation to total chlorophyll. Overall, the chlorophyll concentrations in the SynSenPFT and OC-CCI (Figure 2c) data sets are very similar. However, compared to the OC-CCI data, the SynSenPFT total sum of the three groups shows higher concentrations in the equatorial Pacific around  $40^\circ\text{S}$  and in most parts of the North Atlantic. A further assessment regarding the representation of the sum of the three phytoplankton groups of SynSenPFT (representing total chlorophyll) is discussed in section 3.3.

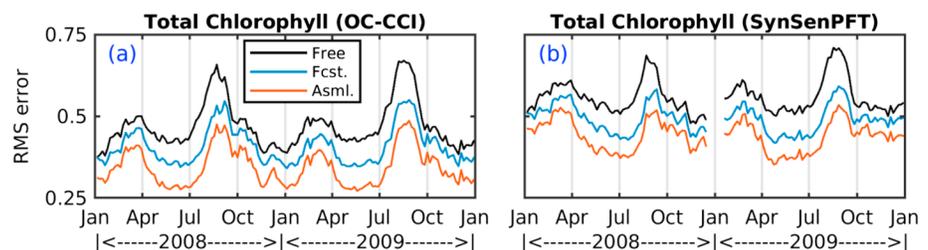
To quantify the global influence of the data assimilation, Figure 3 shows the RMS error of the modeled total surface chlorophyll concentrations with respect to the assimilated OC-CCI (Figure 3a) and the validation data from SynSenPFT (Figure 3b) for the years 2008 and 2009. For statistical consistency, the RMS error (RMSE) is computed from the logarithmic ( $\log_{10}$ ) concentrations in intervals of 5 days, according to the analysis cycles. Shown are the RMSEs for the free run, the 5-day forecasts, and the analysis states.

The analysis estimates (red) show the smallest deviation from the observations, while the free run (black) has the largest errors. The improvement due to the data assimilation ranges between 0.15 and 0.2. The RMSE of the forecast (blue) is consistently higher than that of the analysis showing that the analysis step reduces the deviation from the observations. This is the expected effect of the data assimilation when comparing to the assimilated OC-CCI data. In all three cases, the RMSEs show a seasonal cycle with elevated errors before the spring bloom in both hemispheres and after the spring bloom in the Northern Hemisphere. This behavior is caused by low model concentrations in the high latitudes compared to the SynSenPFT data just before the spring bloom in case of small phytoplankton and after the spring bloom for diatoms.

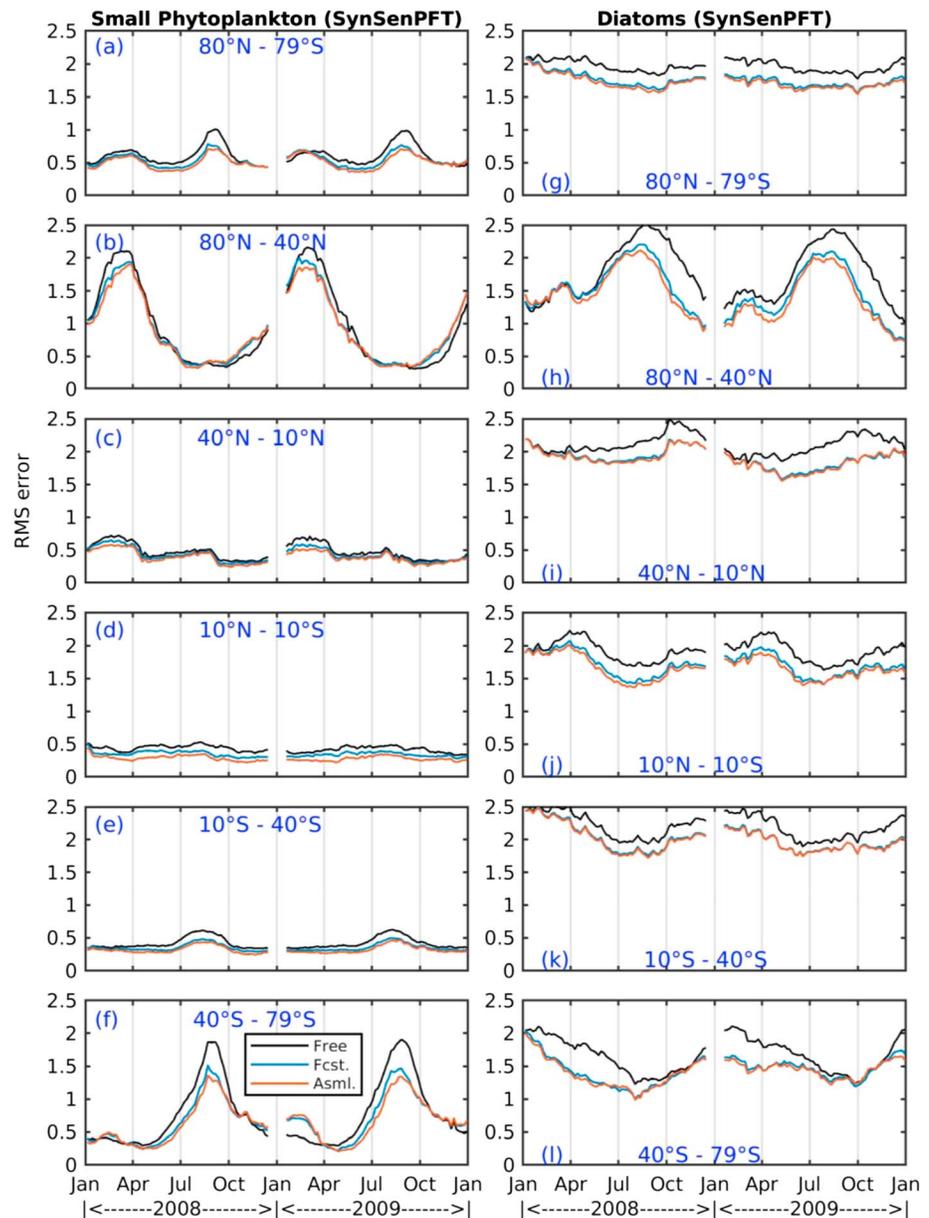
The temporal variation of the errors is similar for the assimilated OC-CCI data and the SynSenPFT data, which represent only three functional groups. However, the RMSEs with regard to the SynSenPFT data



**Figure 2.** Total chlorophyll concentration in milligrams per cubic meter on 20 April 2008 during the spring bloom. (a) Free run, (b) data assimilation analysis state, (c) assimilated OC-CCI data, (d) SynSenPFT data derived from the sum of the chlorophyll concentration of the three groups, (e) difference of free run and OC-CCI data, and (f) difference of assimilation and OC-CCI data. OC-CCI = Ocean-Color Climate Change Initiative.



**Figure 3.** Global RMS errors for total chlorophyll of the model with respect to (a) OC-CCI data and (b) SynSenPFT data. OC-CCI = Ocean-Color Climate Change Initiative. RMS = root-mean-square.



**Figure 4.** RMS error of chlorophyll concentration with regard to SynSenPFT data for individual phytoplankton groups: (a–f) small phytoplankton, (g–l) diatoms in the global domain and five subdomains. RMS = root-mean-square.

are up to 0.1 higher. Overall, the RMSEs show that the data assimilation successfully corrects the total chlorophyll concentrations.

### 3.2. Assessment of Single Phytoplankton Groups With SynSenPFT Data

The SynSenPFT data set provides individual concentrations for the phytoplankton groups of diatoms and some small phytoplankton (coccolithophores plus cyanobacteria). Here we use these data to assess the influence of the data assimilation on the two phytoplankton groups of REcoM2. While the sum of the chlorophyll concentrations of both groups is updated directly by assimilating the total chlorophyll, which is observed by the OC-CCI satellite data, the biomass of the single phytoplankton groups is updated by the multivariate assimilation through the ensemble-estimated cross covariances. Thus, the SynSenPFT data allow us to assess how far this multivariate assimilation is successful. Figure 4 shows RMSEs for the two individual phytoplankton groups for the years 2008 and 2009, for the free model run, the forecast, and the analysis

estimates. Figures 4a and 4g show the RMSEs for the global domain. In general, both groups show improvements due to assimilation. The magnitude of RMSEs is overall higher for diatoms than for the small phytoplankton. This is caused by a bias between SynSenPFT and the model caused by relatively high diatom concentration in the SynSenPFT data, and potentially too low concentrations in the model, in the low-concentration regions of the subtropical gyres (see Losa et al., 2017a).

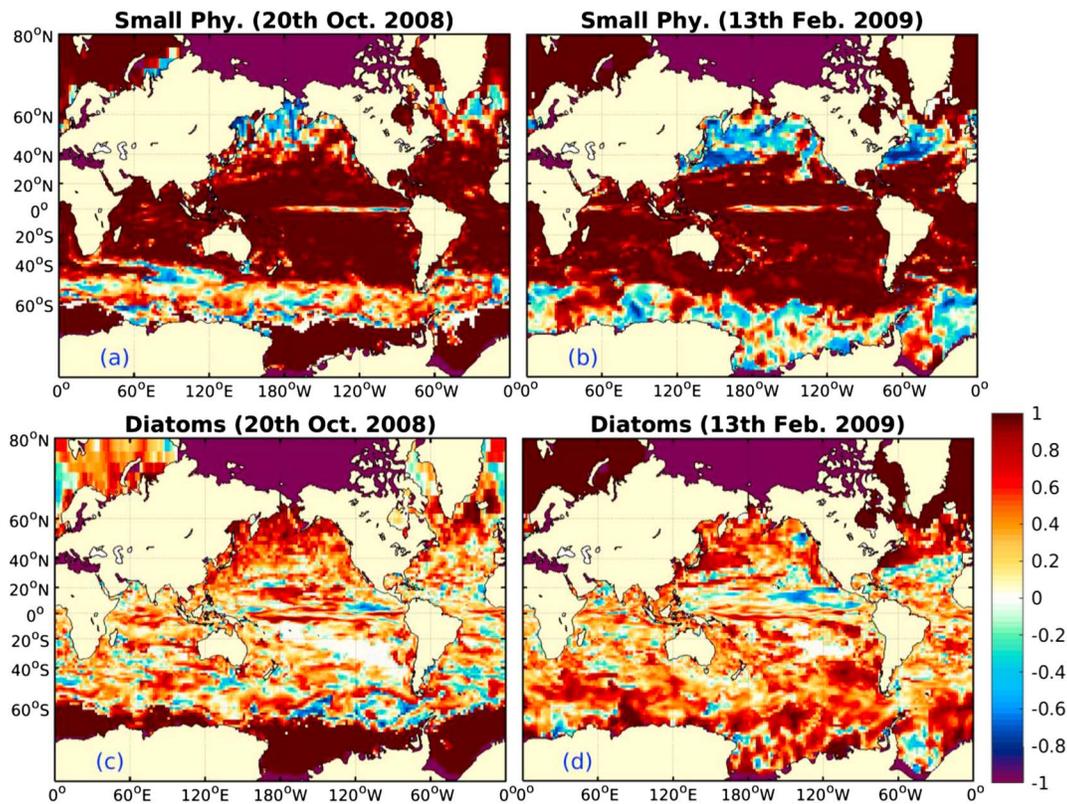
To quantify the regional effect of the data assimilation, we split the global domain into five zonal bands (northern 80–40°N, northern central region 40–10°N, equatorial 10°N–10°S, southern central region 10–40°S, and Antarctic basin 40–79°S). The RMSEs in these bands are shown from the north to south in Figures 4b–4f and Figures 4h–4l. The RMSEs in the central and equatorial regions show low variability of the RMSEs for both small phytoplankton and diatoms. In the higher latitudes in the Northern and Southern Hemispheres a clear seasonal variability is visible for both the groups. However, the variability is distinct for small phytoplankton and diatoms, as the highest RMSEs of the small phytoplankton appear during the spring bloom, while for diatoms the errors are higher in the late summer.

Both Figures 3 and 4 show that the concentration of total chlorophyll and of the phytoplankton groups is improved not only in the analysis state but also in the 5-day forecasts. The forecast deteriorates the analysis fields only slightly. However, the forecast compared to the analysis increases the error by varying magnitudes in different regions, for total chlorophyll and phytoplankton groups; for example, the central regions show smaller error increases than the equatorial region. The error increase in Figure 3 is visually larger than in Figure 4. However, for the total chlorophyll the average error increase is 0.05, while it is 0.032 for small phytoplankton and 0.04 for diatoms (see Tables 1 and 2). Thus, the combined effect of both phytoplankton groups is well in the range of the error increase of total chlorophyll.

The time-mean RMSEs are summarized in Tables 1 and 2. Here the RMSEs are shown for the filter configuration using the cross covariances as well as for the case that the PFTs fields are updated by preserving their ratio to total chlorophyll (rightmost columns). For the small phytoplankton both update variants provide very similar RMSEs with differences up to 1%. In contrast, the RMSEs for the diatoms are about 10% lower when the multivariate assimilation is computed according to the cross covariances compared to the ratio-preserving assimilation update. Thus, using the correlation information between total chlorophyll and the diatoms can significantly improve the estimate of these PFTs.

Figure 4 shows that for most of the time, the data assimilation reduces the RMSEs. However, there are special periods in which the RMSE of at least one of the two groups shows a different behavior. For example, in the Antarctic basin around mid-August 2008, the RMSE is overall high for the small phytoplankton. However, the assimilation results in a significantly lower error than the free run. The situation changes until the beginning of November 2008 when the RMSEs are overall lower than around mid-August 2008, but the error level is the same for both the free run and the assimilation. This behavior repeats again 1 year later. The reasons for this change lie in the beginning of spring in the Southern Hemisphere and in the spatial coverage of the satellite data. In mid-August the chlorophyll concentrations are generally low in the Southern Ocean. The satellite data are available only north of about 50°S. Between 40°S and 50°S satellite data mainly show higher concentrations than the model. The data assimilation can increase these low concentrations, hence reducing the RMSEs. The situation changes in the beginning of November. Now the blooming regions in the model show a significantly higher concentration of chlorophyll from small phytoplankton, and also diatoms, than the satellite data. The effect of the assimilation is now spatially more variable with regions where the deviations from the observations of small phytoplankton are reduced, while they are increased at others. The RMSE shown in Figure 4 is a spatial average. Here the effects of the assimilation average out so that the RMSE is not reduced by the assimilation at this time. This averaging effect is similarly present for the diatoms, so that also here the RMSE around the beginning of November 2008 is not reduced by the data assimilation.

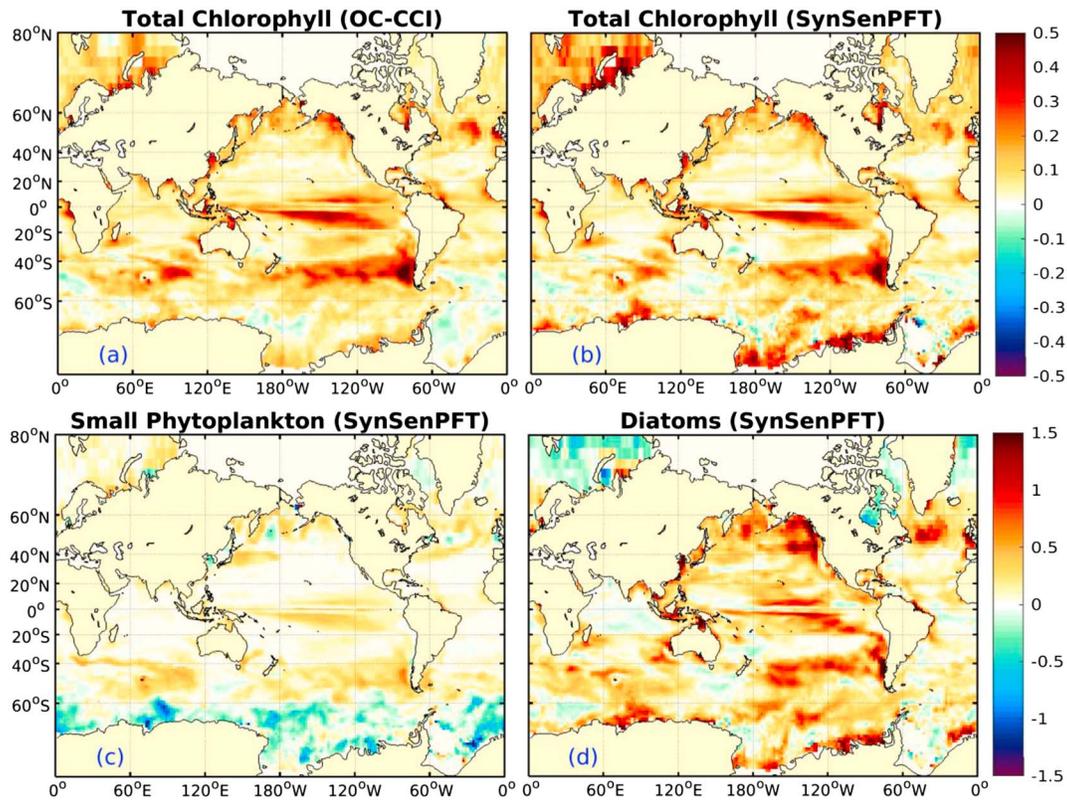
Another particular effect of the assimilation on small phytoplankton is visible in the Antarctic basin around the beginning of February 2009. Here the RMSEs from the assimilation are larger than those from the free run for both the forecast and the analysis fields. In contrast, the diatoms are strongly improved. Actually, in Figure 4 it is visible that the RMSE for the analysis field of small phytoplankton is slightly higher at this time than for the forecast. Thus, the data assimilation increases the error, indicating that the ensemble-estimated cross covariances between total chlorophyll and small phytoplankton are not realistic at this time and



**Figure 5.** Correlation of the model phytoplankton functional types with model total chlorophyll for (a and c) 20 October 2008 and (b and d) 13 February 2009.

location. While the satellite data show higher concentrations than the model, the data assimilation only increases the chlorophyll concentrations of diatoms, while those for small phytoplankton are reduced. In combination, the RMSEs of total chlorophyll are reduced at this time because the effect on the diatoms dominates. A more detailed analysis shows that the concentration of chlorophyll from the small phytoplankton is below  $10^{-3}$  mg/m<sup>3</sup> at many locations where the small phytoplankton is further reduced by the assimilation. Since at the same time the concentration of the diatoms is much larger, the total chlorophyll is only marginally influenced by the reduced concentration of the small phytoplankton, while the increase of diatoms is much more relevant. However, the RMSEs in Figure 4 are computed from the logarithmic concentration. Since this yields relative errors, they look large for the small phytoplankton. The cross correlations for 13 February 2009 between total chlorophyll and the PFTs are shown in Figures 5b and 5d. Here negative correlations are visible for small phytoplankton, for example, around 90°E, 60–65°S, while the correlation is strongly positive for diatoms. This region is one of the places where the small phytoplankton is deteriorated, while diatoms are improved. The unrealistic cross correlations for small phytoplankton appear to be a side effect of the model behavior to enforce a lower concentration limit of  $10^{-4}$  mg/m<sup>3</sup>. If the assimilation update reduces the concentration in an ensemble member below this limit, the model will restore it to the limit. When this modification happens in some ensemble members, it will change the cross covariances so that they are no longer generated by the true model dynamics and can hence be unrealistic.

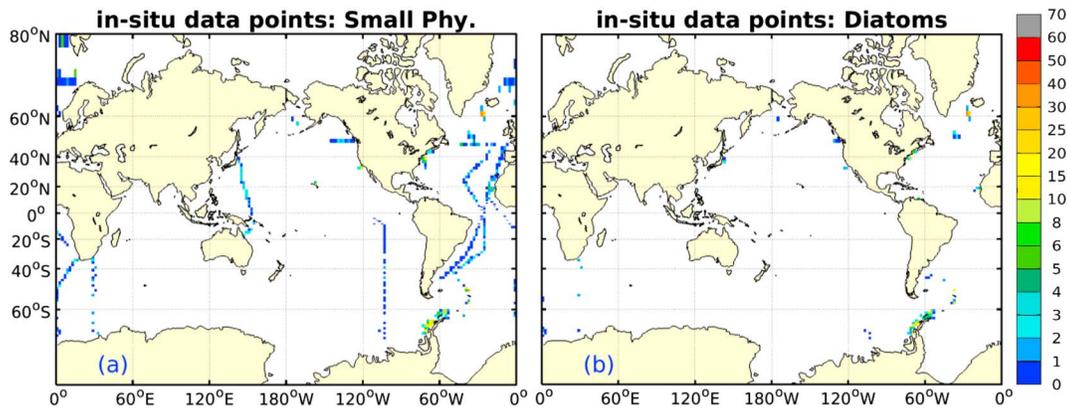
A similar effect in the RMSE is visible in the northern region around the beginning of November 2008. Again, the RMSEs for small phytoplankton for the assimilation are larger than those of the free run and the analysis error is slightly larger than the forecast error. This effect mostly occurs on the North Pacific. Here the model underestimates the concentrations. The chlorophyll from diatoms is strongly improved (i.e., increased) by the data assimilation, while small phytoplankton is decreased in the Bering Sea. As Figure 5a shows, the cross correlation of small phytoplankton with total chlorophyll is again negative, while it is strongly positive for diatoms (Figure 5c).



**Figure 6.** Time-averaged logarithmic assimilation improvement of model chlorophyll. Shown are the improvement of total chlorophyll with regard to OC-CCI data (a), total chlorophyll with regard to SynSenPFT data (b), and the improvement of small phytoplankton (c) and of diatoms (d) with regard to the relating SynSenPFT data. OC-CCI = Ocean-Color Climate Change Initiative.

The time-averaged effect of the data assimilation is shown in Figure 6. Here the quantity “improvement” averaged over the 2 years of the data assimilation experiment is shown. At a given analysis time, the improvement is defined as the absolute deviation of the free run from the observational data set minus the absolute deviation of the assimilation analysis state from the observational data set. The improvement is computed for log<sub>10</sub> concentrations at each analysis time and then averaged over the two years 2008 and 2009. A positive improvement shows that the data assimilation reduces the deviation from the observations, while a negative value shows an increased deviation or deterioration. Since the comparison is made on a log<sub>10</sub> scale, a positive value of one shows that the model estimate with data assimilation is 1 order of magnitude closer to the observations compared to the free run.

The improvement of total chlorophyll with regard to the assimilated OC-CCI data is shown in Figure 6a. Improvements are seen almost everywhere in the global ocean with different amplitude. The largest improvements reaching a value of 0.5 are visible in the equatorial Pacific, in the southern Pacific, and Indian Ocean just south of 40°S. Further significant improvements are visible in both the North Pacific and Atlantic between 40°N and 50°N and in higher latitudes of the Northern Hemisphere. Small deteriorations up to 0.1 are visible in the Atlantic south of 40°S, in a small region east of Australia, and partly in the Pacific south of 60°S. The improvement with regard to SynSenPFT is shown Figure 6b. The improvements with regard to SynSenPFT and OC-CCI data are very similar in most regions. However, the improvements close to the Antarctic and at higher northern latitudes are larger with regard to the SynSenPFT data. Overall, the improvements with regard to SynSenPFT are more variable in the Antarctic basin (south of 40°S) than for OC-CCI data. Several smaller regions with deteriorations of a small amplitude not exceeding 0.1 are visible for SynSenPFT data, and a larger deterioration exists east of the Antarctic peninsula. This also shows that the variability of SynSenPFT data is distinct from that of the OC-CCI data in this region. For both satellite data sets there is a deterioration east of Australia around 40°S, but the amplitude is larger with regard to the SynSenPFT data.



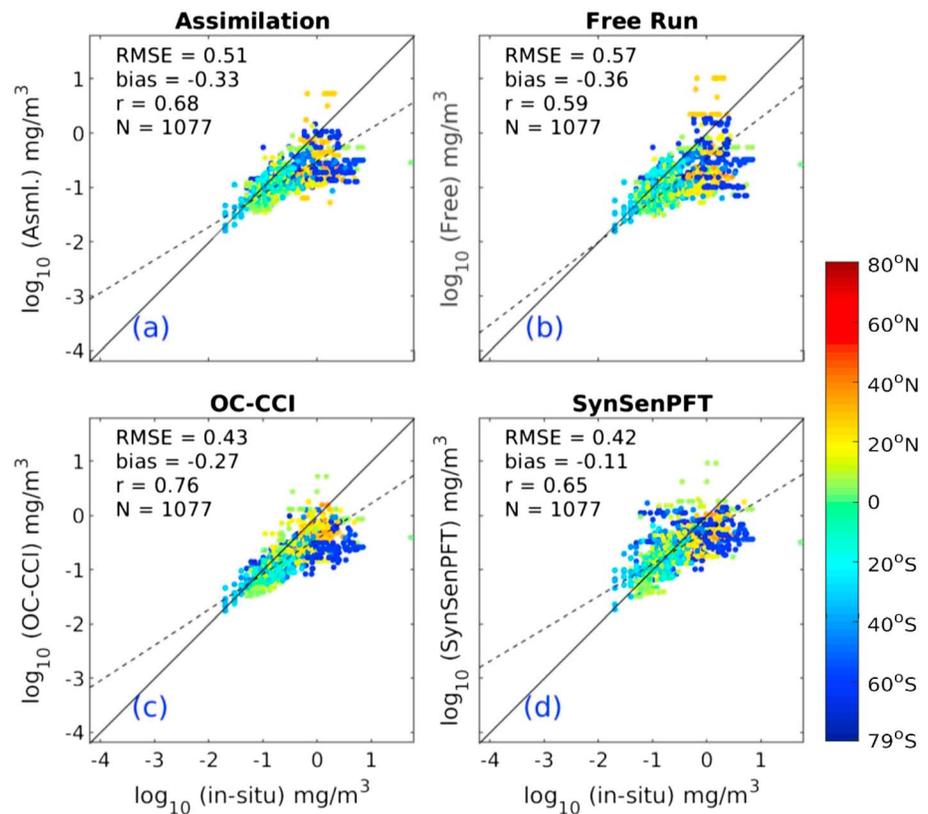
**Figure 7.** Number of in situ data points and their spatial distribution for the years 2008–2009. (a) The 1,054 data points of small phytoplankton and (b) 363 data points of diatoms. The in situ data for total chlorophyll contains 1,077 data points, with approximately the same distribution as the small phytoplankton data.

The assimilation effect on individual phytoplankton groups can be seen in the time-averaged improvement plots for small phytoplankton and diatoms in Figures 6c and 6d which are computed with regard to the phytoplankton groups in the nonassimilated SynSenPFT data. For small phytoplankton, the time-averaged improvement is smaller than for the diatoms. Improvements are visible in most regions like the higher latitudes of the Northern Hemisphere, North Atlantic and Pacific, around the equator, and in temperate regions of the Southern Hemisphere. Deteriorations are mostly present in the Southern Ocean south of 60°S. Here the impact of the assimilation is mainly between neutral and negative, but at some places the data assimilation still improves the state. When we compare the improvement with the RMSEs for this region in Figure 4, we see that most of the time, the data assimilation has a positive effect, while in the Antarctic summer, the small phytoplankton chlorophyll is not improved. The improvements for diatom chlorophyll are generally larger than for the total chlorophyll, in particular in the North Atlantic, North Pacific, equatorial Pacific, and Southern Ocean. This might look contradictory given that total chlorophyll is the sum of the two PFTs, but the shown improvements are relative and not absolute due to the use of logarithmic concentrations. Deteriorations are found in higher latitudes of the Northern Hemisphere and in some smaller regions of the Antarctic. Apart from two very small regions in the Kara Sea and the Hudson Bay with larger deteriorations, the deteriorations are much smaller than the improvements. Thus, the assimilation provides a better state estimate than the free run and provides a complete global coverage in contrast to the satellite data that contain data gaps due to the light availability or clouds.

### 3.3. Assessment of Data Assimilation Influence With In Situ Data

The model results and satellite data are compared against the in situ data by determining the RMSE, bias (mean error), and correlation coefficient based on logarithmic chlorophyll concentrations at the ocean surface. Figure 7 shows the availability of the in situ data over the 2 years. For total chlorophyll there are 1,077 data points, while there are 1,054 data points for small phytoplankton. The data are irregularly distributed over all ocean basins. For diatoms there are only 363 data points and their distribution is very irregular. There are many points close to the Antarctic Peninsula. Also, in the North Atlantic there are several stations with multiple measurements during the 2 years, while the data availability in the Pacific is extremely low and there are essentially no data points in the Indian Ocean.

For total chlorophyll, Figure 8 shows the scatterplots for the assimilation analysis state, the free run, and the OC-CCI and SynSenPFT data compared with in situ data. The color of the points in the scatterplots represents data at different latitudes as shown in the color bar. The correlation ( $r$ ), RMSE, and bias were calculated for  $N = 1,077$  available observations of 2008 and 2009. Note that the total number of points  $N$  is the nearest model grid point to the in situ data. Of the four cases, the OC-CCI data show the highest correlation (0.76) with the in situ data. The free run shows a correlation of 0.59, which is increased by the data assimilation to 0.68. More specifically, the deep blue dots, that is, the total chlorophyll in the southern higher latitudes, are improved. The correlation in the assimilation is slightly higher than the correlation of SynSenPFT data (0.65). For the latter we have to consider that the total sum of all SynSenPFT groups is



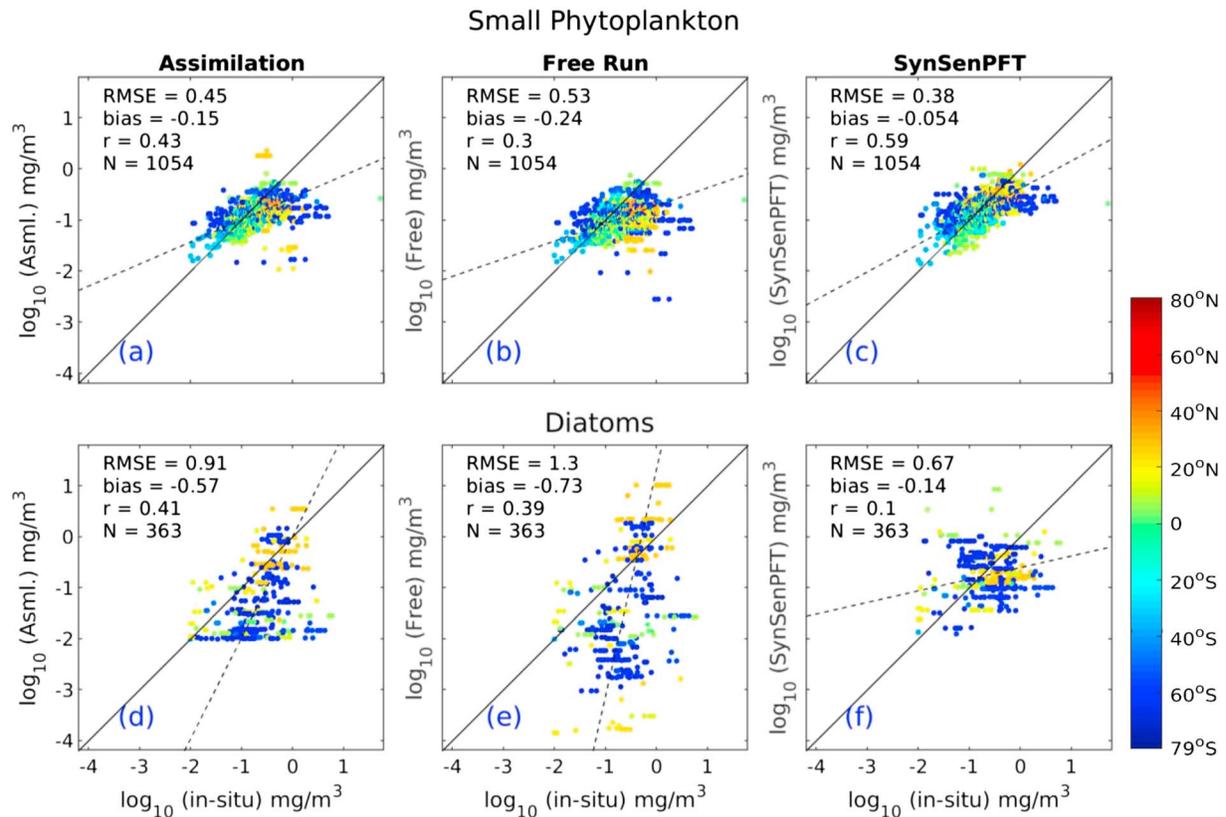
**Figure 8.** Comparison of in situ data for total chlorophyll with the (a) data assimilation analysis, (b) free run, (c) OC-CCI data, and (d) SynSenPFT (see explanation in Figure 2) data. The dashed line shows the linear regression, while the continuous line is the reference line with slope 1. The numbers show the RMSE, bias, correlation coefficient ( $r$ ), and the number of comparison points ( $N$ ). OC-CCI = Ocean-Color Climate Change Initiative; RMSE = root-mean-square.

not identical to the total chlorophyll concentration, because often other phytoplankton groups can contribute significantly (e.g., dinoflagellates, other prymnesiophytes than coccolithophores, pelagophytes, and chrysophytes). Apart from the increase in correlation coefficient the data assimilation reduced the amount of bias (from  $-0.36$  to  $-0.33$ ) and reduced the RMSE (from  $0.57$  to  $0.51$ ). However, both the OC-CCI and SynSenPFT data show smaller biases and RMSEs. So the model state from the data assimilation cannot outcompete these two observational data sets but provide improved coverage for with comparable RMSE and bias.

The data assimilation also has a positive influence on the two individual phytoplankton groups when compared to the in situ data. The scatterplots for small phytoplankton and diatoms are shown in Figure 9. As mentioned in section 2.3.3 only values greater than and equal to  $0.01 \text{ mg/m}^3$  are considered here for both model analysis state and in situ data.

For small phytoplankton (Figures 9a–9c) we have 1,054 data points. The data assimilation reduces the RMSE from  $0.53$  to  $0.45$  and the amplitude of the negative bias from  $-0.24$  to  $-0.15$ . Further, the correlation coefficient is increased from  $0.3$  to  $0.43$  due to the assimilation. Again, the correlation coefficient between SynSenPFT and in situ data is higher with value of  $0.59$ , and the bias and RMSE are lower. This is partly due to the fact that there are a few points with very small concentrations in the model, which are higher in the SynSenPFT data. Comparing the free run and the assimilation, we see that the lowest concentrations are smaller in the free run. These very low concentrations are increased above the limit of  $0.01 \text{ mg/m}^3$  by the data assimilation.

The scatterplots for the diatoms (Figures 9d–9f) show a much larger spread of the values. The number of available in situ points is 363, hence only a third of the data points for small phytoplankton. We see a higher RMSE for diatoms compared to the small phytoplankton, which is reduced from  $1.3$  to  $0.91$  by the data



**Figure 9.** Comparison of in situ data of small phytoplankton (a–c) and diatoms (d–f) with the (a, d) data assimilation analysis state, (b, e) free run, and (c, f) SynSenPFT data. RMSE = root-mean-square.

assimilation. Also, the amount of bias is decreased from  $-0.73$  to  $-0.57$  and the correlation is slightly increased (from  $0.39$  to  $0.41$ ). Like in the total chlorophyll the small concentrations of diatoms in the Southern Ocean are improved (as seen in the deep blue dots). The large improvements in RMSE and bias are mainly caused by a large number of collocation points that are below  $0.01 \text{ mg/m}^3$  in the free run but increased above this limit by the data assimilation. For both the free run and the assimilation, the correlations with the in situ are higher than for the SynSenPFT data which show a very small value of  $0.1$  for the 2-year match-up period but on the other hand a much lower bias and RMSE than the two former data sets ( $-0.14$  vs.  $-0.57$  and  $-0.73$  and  $0.67$  vs.  $0.91$  and  $1.3$ , respectively). Note that the low correlation value for SynSenPFT appears to be particular for the time period of the experiments and due to the low number of match-up points on the model grid. For a 10-year period, 2002 to 2012, Losa et al. (2017a, 2017b) found a better representation with much higher correlation of  $r = 0.67$ , smaller RMSE =  $0.53$ , and bias =  $0.05$  for 4,946 match-up points.

#### 4. Discussion

The results presented above demonstrate that the assimilation of the total chlorophyll from OC-CCI is successful in improving the chlorophyll representation in the REcoM2 model. The total chlorophyll is clearly improved with respect not only to the assimilated satellite data but also to the SynSenPFT data for which the sum of concentrations from three individual PFTs represent the total chlorophyll in our comparisons. South of  $60^\circ\text{S}$ , the assimilation improvement of total chlorophyll is even larger with regard to SynSenPFT than the assimilated OC-CCI data (Figures 6a and 6b).

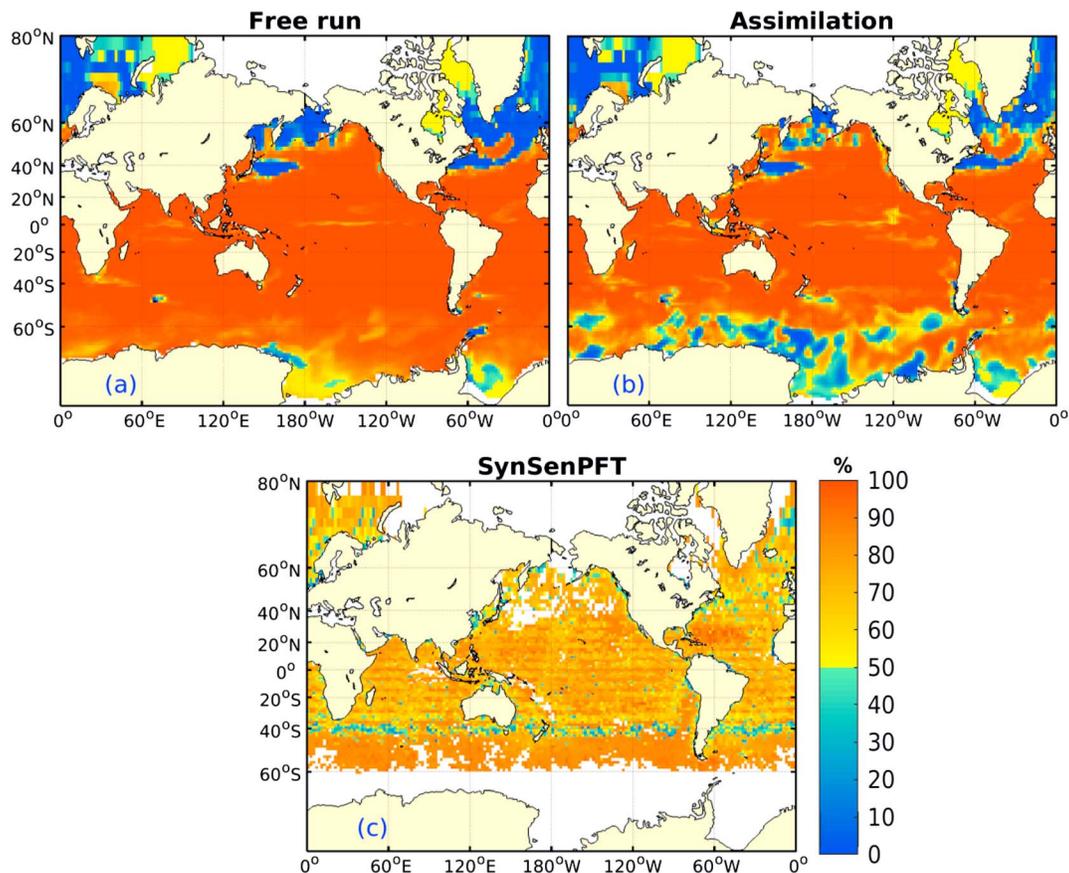
In the South Pacific gyre, a band of increased chlorophyll concentration is visible in Figures 2a and 2b. The assimilation can reduce, but not eliminate, this feature, which is caused by choices in the model parameterization. It is located where a switch from nitrogen to iron limitation happens. The formulation of nutrient

limitation as the minimum of a N- and a Fe-limiting term leads to a small band of less severe limitation at the transition from one to the other, causing the local increase in chlorophyll concentration. A higher model resolution along with fine tuning of the biogeochemistry parameters is expected to reduce this effect.

The improvements reached by data assimilation are partitioned over the two phytoplankton groups represented in the REcoM2 model. Here the ensemble estimates of the cross covariances between total chlorophyll and the chlorophyll in the two phytoplankton groups come into effect. Figure 4 shows that the effect of the data assimilation is to generally reduce the RMSEs for both groups. However, while for diatoms we see strong improvements almost everywhere, the effect is smaller for small phytoplankton. In particular, there are time periods where the assimilation increases the errors for small phytoplankton in higher latitudes. This effect is also visible in Figure 6, where for small phytoplankton south of 60°S both improvements and deteriorations are visible. However, one has to interpret this result with care. While the diatoms have a clear representation in the SynSenPFT data, we approximated the small phytoplankton of REcoM2 by the sum of coccolithophores and cyanobacteria. While REcoM2's small phytoplankton group also includes coccolithophores, there is no exact correspondence; especially the very small cyanobacteria (*Prochlorococcus* and *Synechococcus*) are probably not well represented in the model. Even more as the biogeochemical model parameters are globally constant, the small model phytoplankton cannot represent the spatially varying abundance of different phytoplankton types present in the small phytoplankton. Such an effect has, for example, been discussed by Tréguer et al. (2018) for the diversity within diatoms. Another effect is that the ensemble spread is generated and maintained by perturbing a set of eight model parameters of REcoM2. While previous studies were successful with this strategy and also our results show a mainly positive effect of the assimilation, one has to keep in mind that these parameter perturbations are not tuned to give particularly realistic cross covariances. One might even consider that spatially varying sets of perturbations might be required to improve, for example, the assimilation effect on small phytoplankton in the Antarctic basin.

Tables 1 and 2 show that the multivariate assimilation using the ensemble-estimated cross covariances between total chlorophyll and both PFTs yielded smaller errors for the diatoms than an assimilation configuration in which the PFTs are updated so that the ratio to total chlorophyll is preserved. Further, the results showed that diatoms and small phytoplankton behave differently in the assimilation. To understand this different behavior, Figure 5 exemplifies the ensemble cross correlations between total chlorophyll and both PFTs for two dates. For small phytoplankton (Figures 5a and 5b) the correlation is close to one essentially everywhere between 40°N and 40°S. Only in the higher latitudes smaller, but also negative, correlations are visible. In contrast, the correlation between the diatoms and total chlorophyll is very variable and there are regions, for example, the southern central Pacific where the correlation is close to 0. In the higher latitudes above 40° the correlations show seasonal changes. For example, in the North Atlantic there is a region where small phytoplankton is negatively correlated to total chlorophyll in February, while the correlation is positive in October. In the ratio-preserving assimilation update, the two PFTs are always positively correlated to the total chlorophyll. As such, it is evident that in particular the assimilation effect on the diatoms should be very different for the two update variants. Further, the different correlation patterns of small phytoplankton and diatoms indicate that both PFTs should show distinct assimilation effects when the cross covariances are used for the updates.

Due to the different effect of the assimilation on the two PFTs, the PFT community structure is influenced by the assimilation. Figure 10 shows the fraction of small phytoplankton in the PFT community for the free run, the assimilation experiment, and SynSenPFT for 20 April 2018. Most striking is that the assimilation switches the dominance from small phytoplankton to diatoms in several regions in the Antarctic basin. In contrast, the dominance of diatoms is reduced in the North Atlantic between 40°N and 60°N and to a lesser degree in the North Pacific. For SynSenPFT, small phytoplankton is less dominant in the equatorial and central regions while diatoms show no clear dominance in the northern regions. Around 40°S diatoms dominate, like in the assimilation, but in the data this dominance does not extend further south. The different phytoplankton groups represent different sizes and shapes, but also different physiological characteristics. Further, diatoms and small phytoplankton have different ecological functions like silification versus partial calcification. Changing the phytoplankton community will thus, for example, influence the air-sea flux of CO<sub>2</sub> and particulate organic carbon export (Mouw et al., 2016) and has consequences, for example, for the recruitment of juvenile fish (Trzcinski et al., 2013).



**Figure 10.** Fraction of small phytoplankton in total chlorophyll in percent on 20 April 2018. For values above 50% (yellow to red colors), the small phytoplankton dominates the plankton community, while diatoms dominate for values below 50% (green to blue). (a) Free run, (b) data assimilation analysis state, and (c) SynSenPFT data.

The nutrients and zooplankton were not included in the state vector and thus are updated only via the model dynamics. Accordingly, one cannot expect any systematic effect of the assimilation on these variables. For spatially averaged dissolved inorganic nitrogen and iron, the indirect effect of the assimilation did not exceed 2% at any time in the global average and the overall spatial distribution of the nutrients was preserved. Changes in the total alkalinity were negligible. However, locally the dissolved inorganic nitrogen was increased in the Kara Sea from about 1.5 to 6.0  $\text{mmol}/\text{m}^3$  and dissolved iron from about 2 to 3  $\mu\text{mol}/\text{m}^3$  (not shown). Figure 6 shows that the chlorophyll was significantly improved by the assimilation in the Kara Sea. These changes apparently also induced the changes in the nutrients. Further, the zooplankton concentration increased by about 17% in the global average. These increases were particularly pronounced in the high latitudes, for example, in the Kara Sea, where the zooplankton concentration increased as a reaction on the changed phytoplankton concentration. Generally, the closed northern boundary at 80°N influences the ensemble states and hence the data assimilation in particular in regions with strong water exchanges with the Arctic. However, the assimilation results do not show a clear evidence of this.

The validation with in situ data also confirmed the positive effect of the data assimilation for total chlorophyll and for the two phytoplankton groups. Here one effect of the assimilation is to increase very low phytoplankton concentrations which occur in the free running model. Differences in the distributions of small phytoplankton and diatom concentrations are particularly visible due to the larger spread of diatom concentrations in both the model and the observations. This results in higher RMSEs and biases for diatoms than for small phytoplankton, while the correlation between both data sets is similar for both groups. Nonetheless, the group-based SynSenPFT satellite-derived product still shows better statistics than the assimilation.

Compared to the previous regional study by Ciavatta et al. (2011) we also find a clear improvement of the diatoms. The other groups in the model used by their study (flagellates, dinoflagellates, and picophytoplankton) are not comparable to the small phytoplankton group of REcoM2. In the English Channel modeled by Ciavatta et al. (2011) we see improvements in particular for the diatoms but to a lesser degree also for the small phytoplankton. However, our model configuration has a much coarser resolution than the model configuration used by Ciavatta et al. (2011) and has fewer PFTs. Further, the in situ data points for diatoms were very unevenly distributed and hence mainly representative to a region close to the Antarctic Peninsula and some stations in the Atlantic. However, for diatoms the assimilation shows clear improvements in comparison to SynSenPFT.

## 5. Summary and Conclusions

A local error-subspace transform Kalman filter (LESTKF) has been applied to assimilate satellite data of total chlorophyll into MITgcm-REcoM2 for estimating phytoplankton fields during the years 2008 and 2009. The assimilation was multivariate so that eight variables describing the two PFTs of REcoM2, including the two chlorophyll concentrations, were directly modified by the assimilation through ensemble-estimated covariances. Also, the alternative to update the two PFTs with the constraint to keep their ratio to total chlorophyll constant was tested. The ensemble members were generated by perturbing sensitive biogeochemistry parameters of the ecosystem model.

The assimilation improved the total chlorophyll, represented by the sum of the chlorophyll concentration of REcoM2's two phytoplankton groups. The RMSE for the total chlorophyll for 2008 and 2009 was decreased by the assimilation in comparison with both the assimilated data OC-CCI and the SynSenPFT, which is a semi-independent data set. The RMSE showed a seasonality with larger values during spring at the respective hemispheres specifically in the polar regions. The total chlorophyll was significantly improved over the whole model domain with largest improvements in the equatorial region, the North Atlantic and Pacific between 50°N and 60°N, and in the Pacific and Indian Oceans between 40°S and 50°S. In the Southern Ocean the improvements were larger with regard to the SynSenPFT data than the assimilated OC-CCI data.

The multivariate assimilation was able to improve the two phytoplankton groups individually, which was assessed by comparison with PFTs data from SynSenPFT and in situ data. The improvements were larger in the tropical and midlatitude regions in comparison to high-latitude regions. The Southern Ocean and the region north of 40°N showed a seasonal variability of the RMSE during the spring bloom season in their respective hemispheres. There are times and regions where the influence by the assimilation is clearly distinct for the two phytoplankton groups. In particular, in the Southern Ocean in January and February an error reduction was visible for diatoms which dominate at this time, while the error for small phytoplankton was increased. This effect shows that the ensemble-estimated covariances lead to an individual assimilation effect on the phytoplankton groups, which at times can also lead to a regional deterioration of one group. The individual effect shows that the model dynamics of both groups behave differently resulting in distinct correlations to the total chlorophyll concentration. This behavior is influenced by the parameter perturbations that were applied to generate the ensemble of model states. These perturbations were selected so that a sufficient spread in the PFT variable concentrations was obtained. For example, the perturbations ensure that both the small phytoplankton and diatoms exhibit variability among the different ensemble members during a spring bloom so that the ensemble represents usable covariances between total chlorophyll and the individual PFT variables. However, as a limitation of the method one has to keep in mind that the perturbations were not particularly tuned to obtain optimal assimilation results. To this end there might be other choices for parameter perturbations that lead to a better assimilation performance.

Overall, while the study shows a positive influence of the multivariate assimilation, its assessment is limited by the fact that the SynSenPFT data are not fully independent from the assimilated OC-CCI data. Further, the validation is limited by the data availability and there is a mismatch between the different groups described by the model, the SynSenPFT products, and the in situ PFT data base (see details on this discussion in Bracher, Bouman, et al., 2017). In particular, while the OC-CCI data directly represent total chlorophyll, the assimilation assumes that the sum of the two PFTs of REcoM2 also represent the total chlorophyll, so a likely effect of other PFTs not included in the model is not taken into account. In addition, the small phytoplankton group in REcoM2 is not exactly represented by the available phytoplankton groups in the

SynSenPFT (which are diatoms, cyanobacteria, and coccolithophores) and in situ (which are diatoms, cyanobacteria, and haptophytes) data.

Given the influence of the model parameters, a next step of this work is to estimate them in combination with the concentrations using the data assimilation. This approach will allow for spatially varying parameters, which likely lead to a better representation in particular of the small phytoplankton group of REcoM2. Another approach is to assimilate the PFTs data provided by SynSenPFT or its input data sets PhytoDOAS and OC-PFT to directly influence the PFTs of REcoM2 in a similar manner as done by Ciavatta et al. (2018) for the northwest European Shelf.

### Acknowledgments

We acknowledge the North-German Supercomputing Alliance (HLRN) for providing HPC resources that have contributed to the research results reported in this paper. This work is funded by the AWI Strategy Fund by Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI), Germany. We acknowledge Euromarine network for funding partially to attend the “Ensemble data assimilation workshop 2016” in Bergen, Norway, which gave H. K. P. deeper insight regarding assimilation. The phytoplankton groups product SynSenPFT data and the in situ data for validation are available at <https://doi.pangaea.de/10.1594/PANGAEA.875873> and <https://doi.pangaea.de/10.1594/PANGAEA.875879>, respectively. S. L. gratefully acknowledges the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 268020496—TRR 172, within the Transregional Collaborative Research Center “Arctic Amplification: Climate Relevant Atmospheric and Surface Processes, and Feedback Mechanisms (AC<sup>3</sup>)” and partly made in the framework of the state assignment of the Federal Agency for Scientific Organizations (FASO) Russia (theme 0149-2019-0015). Finally, the authors would like to acknowledge the two anonymous reviewers for their constructive comments.

### References

- Aumont, O., Maier-Reimer, E., Blain, S., & Monfray, P. (2003). An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles*, *17*(2), 1060. <https://doi.org/10.1029/2001GB001745>
- Bracher, A., Bouman, H. A., Brewin, R. J. W., Bricaud, A., Brotas, V., Ciotti, A. M., et al. (2017). Obtaining phytoplankton diversity from ocean color: A scientific roadmap for future development. *Frontiers in Marine Science*, *4*, 55. <https://doi.org/10.3389/fmars.2017.00055>
- Bracher, A., Vountas, M., Dinter, T., Burrows, J. P., Röttgers, R., & Peeken, I. (2009). Quantitative observation of cyanobacteria and diatoms from space using PhytoDOAS on SCIAMACHY data. *Biogeosciences*, *6*(5), 751–764. <https://doi.org/10.5194/bg-6-751-2009>
- Bracher, A., Dinter, T., Wolanin, A., Rozanov, V. V., Losa, S., & Soppa, M. A. (2017). Global monthly mean chlorophyll “a” surface concentrations from August 2002 to April 2012 for diatoms, coccolithophores and cyanobacteria from PhytoDOAS algorithm version 3.3 applied to SCIAMACHY data, link to NetCDF files in ZIP archive. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.870486>
- Brewin, R. J. W., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Ains, R., & Lamont, T. (2015). Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. *Remote Sensing of Environment*, *168*, 437–450. <https://doi.org/10.1016/j.rse.2015.07.004>
- Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research*, *100*(C7), 13237. <https://doi.org/10.1029/95JC00458>
- Chen, Z., Liu, J., Song, M., Yang, Q., & Xu, S. (2017). Impacts of assimilating satellite sea ice concentration and thickness on Arctic sea ice prediction in the NCEP climate forecast system. *Journal of Climate*, *30*(21), 8429–8446. <https://doi.org/10.1175/JCLI-D-17-0093.1>
- Ciavatta, S., Brewin, R. J. W., Skákala, J., Polimene, L., de Mora, L., Artioli, Y., & Allen, J. I. (2018). Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations. *Journal of Geophysical Research: Oceans*, *123*, 834–854. <https://doi.org/10.1002/2017JC013490>
- Ciavatta, S., Kay, S., Saux-Picart, S., Butenschön, M., & Allen, J. I. (2016). Decadal reanalysis of biogeochemical indicators and fluxes in the north west European shelf-sea ecosystem. *Journal of Geophysical Research: Oceans*, *121*, 1824–1845. <https://doi.org/10.1002/2015JC011496>
- Ciavatta, S., Torres, R., Saux-Picart, S., & Allen, J. I. (2011). Can ocean color assimilation improve biogeochemical hindcasts in shelf seas? *Journal of Geophysical Research*, *116*, C12043. <https://doi.org/10.1029/2011JC007219>
- Doron, M., Brasseur, P., Brankart, J. M., Losa, S. N., & Melet, A. (2013). Stochastic estimation of biogeochemical parameters from GlobColour ocean colour satellite data in a North Atlantic 3D ocean coupled physical-biogeochemical model. *Journal of Marine Systems*, *117–118*, 81–95. <https://doi.org/10.1016/j.jmarsys.2013.02.007>
- Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(C5), 10143. <https://doi.org/10.1029/94JC00572>
- Ford, D., & Barciela, R. (2017). Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products. *Remote Sensing of Environment*, *203*, 40–54. <https://doi.org/10.1016/j.rse.2017.03.040>
- Ford, D. A., Edwards, K. P., Lea, D., Barciela, R. M., Martin, M. J., & Demaria, J. (2012). Assimilating GlobColour ocean colour data into a pre-operational physical-biogeochemical model. *Ocean Science*, *8*(5), 751–771. <https://doi.org/10.5194/os-8-751-2012>
- Friedlingstein, P., Bopp, L., Ciais, P., Dufresne, J.-L., Fairhead, L., LeTreut, H., et al. (2001). Positive feedback between future climate change and the carbon cycle. *Geophysical Research Letters*, *28*(8), 1543–1546. <https://doi.org/10.1029/2000GL012015>
- García, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K., & Johnson, D. R. (2010). World ocean atlas 2009, volume 4: Nutrients (phosphate, nitrate, and silicate). *NOAA World Ocean Atlas*, *119*(1), 227–237. <https://doi.org/10.1182/blood-2011-06-357442>
- Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1998). A dynamic regulatory model of phytoplankton acclimation to light, nutrients, and temperature. *Limnology and Oceanography*, *43*(4), 679–694. <https://doi.org/10.4319/lo.1998.43.4.0679>
- Grant, M., Jackson, T., Chuprin, A., Sathyendranath, S., Zühlke, M., Storm, T., et al. (2015). OC-CCI v2.0 product user guide 2.0.5, 38.
- Gregg, W. W. (2008). Assimilation of SeaWiFS ocean chlorophyll data into a three-dimensional global ocean model. *Journal of Marine Systems*, *69*(3–4), 205–225. <https://doi.org/10.1016/j.jmarsys.2006.02.015>
- Gregg, W. W., & Rousseaux, C. S. (2014). Decadal trends in global pelagic ocean chlorophyll: A new assessment integrating multiple satellites, in situ data, and models. *Journal of Geophysical Research: Oceans*, *119*, 5921–5933. <https://doi.org/10.1002/2014JC010158>
- Hammond, M. L., Beaulieu, C., Sahu, S. K., & Henson, S. A. (2017). Assessing trends and uncertainties in satellite-era ocean chlorophyll using space-time modeling. *Global Biogeochemical Cycles*, *31*, 1103–1117. <https://doi.org/10.1002/2016GB005600>
- Hauck, J., Köhler, P., Wolf-Gladrow, D., & Völker, C. (2016). Iron fertilisation and century-scale effects of open ocean dissolution of olivine in a simulated CO<sub>2</sub> removal experiment. *Environmental Research Letters*, *11*(2), 024007. <https://doi.org/10.1088/1748-9326/11/2/024007>
- Hauck, J., & Völker, C. (2015). Rising atmospheric CO<sub>2</sub> leads to large impact of biology on Southern Ocean CO<sub>2</sub> uptake via changes of the Revelle factor. *Geophysical Research Letters*, *42*, 1459–1464. <https://doi.org/10.1002/2015GL063070>
- Hauck, J., Völker, C., Wang, T., Hoppema, M., Losch, M., & Wolf-Gladrow, D. A. (2013). Seasonally different carbon flux changes in the Southern Ocean in response to the southern annular mode. *Global Biogeochemical Cycles*, *27*, 1236–1245. <https://doi.org/10.1002/2013GB004600>
- Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., et al. (2011). Synoptic relationships between surface chlorophyll-*a* and diagnostic pigments specific to phytoplankton functional types. *Biogeosciences*, *8*(2), 311–327. <https://doi.org/10.5194/bg-8-311-2011>

- Hohn, S. (2009). Coupling and decoupling of biogeochemical cycles in marine ecosystems. *Ecological Modelling*, 135. Retrieved from <http://elib.suub.uni-bremen.de/peid=D00011278>
- Hu, J., Fennel, K., Mattern, J. P., & Wilkin, J. (2012). Data assimilation with a local ensemble Kalman filter applied to a three-dimensional biological model of the Middle Atlantic Bight. *Journal of Marine Systems*, 94, 145–156. <https://doi.org/10.1016/j.jmarsys.2011.11.016>
- Jackson, T., Sathyendranath, S., & Mélin, F. (2017). An improved optical classification scheme for the ocean colour essential climate variable and its applications. *Remote Sensing of Environment*, 203, 152–161. <https://doi.org/10.1016/j.rse.2017.03.036>
- Jones, E. M., Baird, M. E., Mongin, M., Parslow, J., Skerratt, J., Lovell, J., et al. (2016). Use of remote-sensing reflectance to constrain a data assimilating marine biogeochemical model of the Great Barrier Reef. *Biogeosciences*, 13(23), 6441–6469. <https://doi.org/10.5194/bg-13-6441-2016>
- Key, R. M., Kozyr, A., Sabine, C. L., Lee, K., Wanninkhof, R., Bullister, J. L., et al. (2004). A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles* <https://doi.org/10.1029/2004GB002247>, 18, GB4031.
- Large, G., & Yeager, S. (2004). *Diurnal to Decadal Global Forcing for Oceans and Sea-Ice Models. The Data Sets and Flux Climatologies* (NCAR/TN-460+STR, Technical report). Boulder, CO: National Center for Atmospheric Research.
- Levitus, S., Locarnini, R. A., Boyer, T. P., Mishonov, A. V., Antonov, J. I., Garcia, H. E., et al. (2010). World ocean atlas 2009. Retrieved from <https://repository.library.noaa.gov/view/noaa/1259>
- Losa, S. N., Soppa, M. A., Dinter, T., Wolanin, A., Brewin, R. J. W., Bricaud, A., et al. (2017a). Synergistic exploitation of hyper- and multi-spectral precursor sentinel measurements to determine phytoplankton functional types (SynSenPFT). *Frontiers in Marine Science*, 4, 203. <https://doi.org/10.3389/fmars.2017.00203>
- Losa, Svetlana, Soppa, Mariana A., Dinter, Tilman, Wolanin, Aleksandra, Brewin, Robert J W, Bricaud, Annick, et al. (2017b). Global data sets of chlorophyll “a” concentration for diatoms, coccolithophores (haptophytes) and cyanobacteria obtained from in situ observations and satellite retrievals. <https://doi.org/10.1594/PANGAEA.873210>
- Losch, M., Schröder, M., Hohn, S., & Völker, C. (2008). *High-resolution modelling of phytoplankton distribution and adaptation*, John von Neumann Institute for Computing Symposium (Vol. 39, pp. 289–296). Jülich: Forschungszentrum Jülich. Retrieved from <http://www.fz-juelich.de/nic-series/volume39>
- Mahowald, N. (2003). Interannual variability in atmospheric mineral aerosols from a 22-year model simulation and observational data. *Journal of Geophysical Research*, 108(D12), 4352 <https://doi.org/10.1029/2002JD002821>
- Maritorea, S., d’Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, 114(8), 1791–1804. <https://doi.org/10.1016/j.rse.2010.04.002>
- Marshall, J., Adcroft, A., Hill, C., Perelman, L., & Heisey, C. (1997). A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research*, 102(C3), 5753–5766. <https://doi.org/10.1029/96JC02775>
- MITgcm Group. (2018). MITgcm user manual. Retrieved June 13, 2018, from [http://mitgcm.org/public/r2\\_manual/latest/online\\_documents/manual.html](http://mitgcm.org/public/r2_manual/latest/online_documents/manual.html)
- Mouw, C. B., Barnett, A., McKinley, G. A., Gloege, L., & Pilcher, D. (2016). Phytoplankton size impact on export flux in the global ocean. *Global Biogeochemical Cycles*, 30, 1542–1562. <https://doi.org/10.1002/2015GB005355>
- Natvik, L.-J., & Evensen, G. (2003). Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments. *Journal of Marine Systems*, 40–41, 127–153. [https://doi.org/10.1016/S0924-7963\(03\)00016-2](https://doi.org/10.1016/S0924-7963(03)00016-2)
- Nerger, L., Danilov, S., Kivman, G., Hiller, W., & Schröter, J. (2007). Data assimilation with the ensemble Kalman filter and the SEIK filter applied to a finite element model of the North Atlantic. *Journal of Marine Systems*, 65(1–4), 288–298. <https://doi.org/10.1016/j.jmarsys.2005.06.009>
- Nerger, L., & Gregg, W. W. (2007). Assimilation of SeaWiFS data into a global ocean biogeochemical model using a local SEIK filter. *Journal of Marine Systems*, 68(1–2), 237–254. <https://doi.org/10.1016/j.jmarsys.2006.11.009>
- Nerger, L., & Gregg, W. W. (2008). Improving assimilation of SeaWiFS data by the application of bias correction with a local SEIK filter. *Journal of Marine Systems*, 73(1–2), 87. <https://doi.org/10.1016/j.jmarsys.2007.09.007-102>
- Nerger, L., & Hiller, W. (2013). Computers & geosciences software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers and Geosciences*, 55, 110–118. <https://doi.org/10.1016/j.cageo.2012.03.026>
- Nerger, L., Janjić, T., Schröter, J., & Hiller, W. (2012). A regulated localization scheme for ensemble-based Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, 138(664), 802–812. <https://doi.org/10.1002/qj.945>
- Ourmières, Y., Brasseur, P., Lévy, M., Brankart, J. M., & Verron, J. (2009). On the key role of nutrient data to constrain a coupled physical-biogeochemical assimilative model of the North Atlantic Ocean. *Journal of Marine Systems*, 75(1–2), 100–115. <https://doi.org/10.1016/j.jmarsys.2008.08.003>
- Rousseaux, C. S., & Gregg, W. W. (2012). Climate variability and phytoplankton composition in the Pacific Ocean. *Journal of Geophysical Research*, 117, C10006. <https://doi.org/10.1029/2012JC008083>
- Rousseaux, C. S., & Gregg, W. W. (2015). Recent decadal trends in global phytoplankton composition. *Global Biogeochemical Cycles*, 29, 1674–1688. <https://doi.org/10.1002/2015GB005139>
- Sadeghi, A., Dinter, T., Vountas, M., Taylor, B. B., Altenburg-Soppa, M., Peeken, I., & Bracher, A. (2012). Improvement to the PhytoDOAS method for identification of coccolithophores using hyper-spectral satellite data. *Ocean Science*, 8(6), 1055–1070. <https://doi.org/10.5194/os-8-1055-2012>
- Sathyendranath, S., Grant, M., Brewin, R.J.W., Brockmann, C., Brotas, V., Chuprin, A., et al. (2018). ESA Ocean Colour Climate Change Initiative (Ocean\_Colour\_cci): Version 3.1 data. <https://doi.org/10.5285/9c334fbe6d424a708cf3c4cf0c6a53f5>
- Schartau, M., Engel, A., Schröter, J., Thoms, S., Völker, C., & Wolf-Gladrow, D. (2007). Modelling carbon overconsumption and the formation of extracellular particulate organic carbon. *Biogeosciences*, 4(4), 433–454. Retrieved from [www.biogeosciences.net/4/433/2007/](http://www.biogeosciences.net/4/433/2007/). <https://doi.org/10.5194/bg-4-433-2007>
- Shulman, I., Frolov, S., Anderson, S., Penta, B., Gould, R., Sakalaukus, P., & Ladner, S. (2013). Impact of bio-optical data assimilation on short-term coupled physical. *Bio-Optical Model Predictions*, 118(4), 2215–2230. <https://doi.org/10.1002/jgrc.20177>
- Simon, E., Samuelsen, A., Bertino, L., & Mouisset, S. (2015). Experiences in multiyear combined state-parameter estimation with an ecosystem model of the North Atlantic and Arctic Oceans using the ensemble Kalman filter. *Journal of Marine Systems*, 152, 1–17. <https://doi.org/10.1016/j.jmarsys.2015.07.004>
- Skákala, J., Ford, D., Brewin, R. J. W., McEwan, R., Kay, S., Taylor, B., et al. (2018). The assimilation of phytoplankton functional types for operational forecasting in the northwest European Shelf. *Journal of Geophysical Research: Oceans*. <https://doi.org/10.1029/2018JC014153>, 123, 5230–5247.
- Soppa, M., Losa, S., Brewin, R., Bricaud, A., & Bracher, A. (2016). SY-4Sci synergy R & D study 4: Phytoplankton functional types (SynSenPFT) product validation report (PVR).

- Soppa, M. A., Hirata, T., Silva, B., Dinter, T., Peeken, I., Wiegmann, S., & Bracher, A. (2014). Global retrieval of diatom abundance based on phytoplankton pigments and satellite data. *Remote Sensing*, 6(10), 10,089–10,106. <https://doi.org/10.3390/rs61010089>
- Soppa, M. A., Peeken, I., & Bracher, A. (2017). Global chlorophyll “a” concentrations for diatoms, haptophytes and prokaryotes obtained with the diagnostic pigment analysis of HPLC data compiled from several databases and individual cruises. *PANGAEA*. <https://doi.org/10.1594/PANGAEA.875879>
- Tjiputra, J. F., Polzin, D., & Winguth, A. M. E. (2007). Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization. *Global Biogeochemical Cycles*, 21, GB1001. <https://doi.org/10.1029/2006GB002745>
- Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O., et al. (2018). Influence of diatom diversity on the ocean biological carbon pump. *Nature Geoscience*, 11(1), 27–37. <https://doi.org/10.1038/s41561-017-0028-x>
- Triantafyllou, G., Korres, G., Hoteit, I., Petihakis, G., & Banks, A. C. (2007). Assimilation of ocean colour data into a biogeochemical flux model of the eastern Mediterranean Sea. *Ocean Science*, 3(3), 397–410. <https://doi.org/10.5194/os-3-397-2007>
- Trzcinski, M., Devred, E., Platt, T., & Sathyendranath, S. (2013). Variation in ocean colour may help predict cod and haddock recruitment. *Marine Ecology Progress Series*, 491, 187–197. <https://doi.org/10.3354/meps10451>
- Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *Journal of Geophysical Research*, 111, C08005. <https://doi.org/10.1029/2005JC003207>
- Valente, A., Sathyendranath, S., Brotas, V., Groom, S., Grant, M., Taberner, M., et al. (2016). A compilation of global bio-optical in situ data for ocean-colour satellite applications. *Earth System Science Data*, 8(1), 235–252. <https://doi.org/10.5194/essd-8-235-2016>
- Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., & Marty, J.-C. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. *Journal of Geophysical Research*, 106(C9), 19,939–19,956. <https://doi.org/10.1029/1999JC000308>
- Xiao, Y., & Friedrichs, M. A. M. (2014). The assimilation of satellite-derived data into a one-dimensional lower trophic level marine ecosystem model. *Journal of Geophysical Research: Oceans*, 119, 2691–2712. <https://doi.org/10.1002/2013JC009433>