CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Plymouth Marine Science Electronic Archive (PlyMSEA)

Improved representation of underwater light field and its impact on ecosystem dynamics: a study in the North Sea

Jozef Skákala[1,2], Jorn Bruggeman[1], Robert J.W. Brewin[3], David A. Ford[4]and Stefano Ciavatta[1,2]

[1]Plymouth Marine Laboratory, The Hoe, Plymouth, PL1 3DH United Kingdom.
[2]National Centre for Earth Observation, Plymouth, PL1 3DH, UK.
[3]College of Life and Environmental Sciences, University of Exeter, Penryn Campus Cornwall, TR10 9EZ, UK.
[4]Met Office, FitzRoy Road, Exeter, EX1 3PB UK.

Key Points:

- We provided a state-of-the-art marine ecosystem model with spectrally and directionally resolved incoming radiation.
- We improved the model skill in representing the underwater light field in the North Sea.
- With the help of assimilation we also improved the model skill in representing ecosystem dynamics.

Corresponding author: Jozef Skákala, jos@pml.ac.uk

Abstract

Understanding ecosystem state on the North-West European (NWE) Shelf is of major importance for both economy and climate research. The purpose of this work is to advance our modelling of in-water optics on the NWE Shelf, with important implications for how we model primary productivity, as well as for assimilation of water-leaving radiances. We implement a stand-alone bio-optical module into the existing coupled physical-biogeochemical model configuration. The advantage of the bio-optical module, when compared to the pre-existing light scheme is that it resolves the underwater light spectrally and distinguishes between direct and diffuse downwelling streams. The changed underwater light compares better with both satellite and in-situ observations. The module lowered the underwater Photosynthetically Active Radiation, decreasing the simulated primary productivity, but overall the improved underwater light had relatively limited impact on the phytoplankton seasonal dynamics. We showed that the model skill in representing phytoplankton seasonal cycle (e.g phytoplankton bloom) can be substantially improved either by assimilation of satellite Phytoplankton Functional Type (PFT) chlorophyll, or by assimilating a novel PFT absorption product. Assimilation of the two PFT products yields similar results, with an important difference in the PFT community structure. Both assimilative runs lead to lower plankton biomass and increase the nutrient concentrations. We discuss some future directions on how to improve our model skill in biogeochemistry without using assimilation, e.g. by improving nutrient forcing, re-tuning the model parameters and using the bio-optical module to provide a two-way physical-biogeochemical coupling, improving the consistency between model physical and biogeochemical components.

1 Introduction

Ecosystems convert inorganic matter into organic compounds mostly through the process of photosynthesis. The central role of light in photosynthesis implies that any successful marine or terrestrial ecosystem model must be reasonably skilled in representing the basic properties of the incoming light field. Representation of light is of special importance to marine ecosystem models, since the ocean has a large impact on light properties, pathways, and extinction, mostly through backscattering and absorption by water, phytoplankton, sediment, detritus and particulate organic matter (e.g *Gregg and Casey* [2009]; *Gregg and Rousseaux* [2016, 2017]). What exactly happens with light as it penetrates through the ocean and how much of it is used to drive photosynthesis, depends largely on its spectral and directional properties at the time when light enters the water column. These spectral and directional properties in turn depend on the atmospheric conditions, in particular on the scattering and absorption by clouds, aerosols, ozone, water vapour and other atmospheric constituents (*Gregg and Casey* [2009]; *Gregg and Rousseaux* [2016]).

Although biological processes depend on the light spectral decomposition (*Dickey et al.* [2011]), most ecosystem models represent the underwater irradiance either by a total short-wave radiation, or Photosynthetically Active Radiation (PAR) (*Palmer and Totterdell* [2001]; *Zielinski et al.* [2002]; *Maier-Reimer et al.* [2005]; *Doney et al.* [2006]; *Henson et al.* [2010]; *Marinov et al.* [2010]; *Laufkötter et al.* [2013]; *Butenschön et al.* [2016], see also *Gregg and Rousseaux* [2016] for an overview). Furthermore, the direction of the light beam remains often unresolved, which typically equals to the assumption that the incoming radiation is 100% diffuse (e.g *Butenschön et al.* [2016]). The underwater light field is in those ecosystem models reduced through a broadband attenuation term (e.g *Zhao et al.* [2013]; *Butenschön et al.* [2016]), which sometimes distinguishes between the roles of water and phytoplankton in light attenuation (*Jiang et al.* [2003]; *Manizza et al.* [2005]; *Xiu and Chai* [2014]). However, in the last two decades there have appeared a number of one-dimensional (1D, *Bissett et al.* [1999, 2005]) and three-dimensional (3D, *Gregg* [2002]; *Gregg et al.* [2003]; *Gregg and Casey* [2007]; *Mobley et al.* [2009]; *Dutkiewicz et al.* [2015]; *Baird et al.* [2016]; *Gregg and Rousseaux* [2016, 2017]) modelling stud-

ies where light has been spectrally resolved. For example, a recent study by *Gregg and Rousseaux* [2016] highlighted the importance of spectrally and directionally resolved light to simulate the global phytoplankton community structure, as well as the global chlorophyll and nutrient abundances. *Gregg and Rousseaux* [2016] have also shown that global primary productivity can be highly sensitive (on the levels of an order of magnitude) to the wavelength chosen to represent the broadband radiation in a non-spectra resolving light model.

The seas covering continental shelf have often nutrient-rich, biologically productive waters (*de Haas et al.* [2002]), due to high levels of mixing in their shallow bathymetry and river input. North-West European (NWE) Shelf is a region of natural interest to European fisheries and economy (*Pauly et al.* [2002]), and it is also a region of major importance for the carbon cycle and climate (e.g *Borges et al.* [2006]; *Jahnke* [2010]). The operational model run on the NWE Shelf consists of physical model Nucleus for European Modelling of Ocean (NEMO) coupled through the Framework for Aquatic Biogeochemical Model (FABM, *Bruggeman and Bolding* [2014]) to the biogeochemical model the European Regional Seas Ecosystem Model (ERSEM). ERSEM is a popular lower trophic level model to represent Shelf Sea biogeochemistry (*Baretta et al.* [1995]; *Blackford and Gilbert* [2007]; *Holt et al.* [2012]; *Wakelin et al.* [2012]; *Polimene et al.* [2012]; *Artioli et al.* [2012]; *Butenschön et al.* [2016]), which is the biogeochemical component of the operational model for the NWE Shelf used by the European Copernicus Marine Ecosystem Monitoring Service (CMEMS). ERSEM skill has been repeatedly validated against different types of data (*Allen and Somerfield* [2009]; *Edwards et al.* [2012]; *De Mora et al.* [2013, 2016]). However, ERSEM suffers from the same limitations as most marine ecosystem models: in the established ERSEM configuration (*Butenschön et al.* [2016]) the light is taken as purely diffuse radiation (approximation for the higher latitudes) and is spectrally unresolved. The ERSEM model is typically forced by an external atmospheric product for net downwelling short-wave radiation, whilst underwater light extinction is calculated from light attenuation by clear sea water, 4 different Phytoplankton Functional Types (PFTs) and from aggregate absorption by Particulate Organic Matter (POM), Colored Dissolved Organic Matter (CDOM) and sediment forced by an external product (*Butenschön et al.* [2016]). The oversimplified ERSEM light scheme poses limitation on how model represents primary productivity (e.g concentrations of phytoplankton biomass, magnitude and timing of phytoplankton bloom). Furthermore, the ERSEM light scheme is not reliable enough to be used in how we calculate the heating of the water column and a separate scheme from the physical model needs to be used instead. These drawbacks in how ERSEM treats underwater light are particularly concerning on the NWE Shelf, given that the Shelf seems to be one of the global regions with higher sensitivity to the representation of in-water optics (*Gregg and Rousseaux* [2016]).

In this work we aim to improve the ERSEM representation of underwater light by implementing a spectrally resolved stand-alone bio-optical module (developed in the context of FABM) into an established NEMO - FABM - ERSEM configuration on the NWE Shelf. In the context of ERSEM, similar efforts have been made in *Ciavatta et al.* [2014], but this study goes far beyond of what has been done in the early paper of *Ciavatta et al.* [2014]. Unlike *Ciavatta et al.* [2014]: a) we distinguish between direct and diffuse downwelling streams; b) we comprehensively resolve light in a broad range of wavelengths including ultraviolet and infrared, while *Ciavatta et al.* [2014] resolved only 3 bands, all in a visible range; c) our model has broad application on the NWE Shelf and beyond, whereas the applicability of the model by *Ciavatta et al.* [2014] was constrained to the Western English Channel. The bio-optical module implemented in this study is based on OceanâĂŞAtmosphere Spectral Irradiance Model (OASIM) (e.g *Gregg and Casey* [2009]) and forced by atmospheric fields, such as ozone, cloud cover and water vapour. It provides ERSEM with spectrally resolved radiation at the water surface and also with improved directional representation of light by decomposing the downwelling radiation into diffuse and direct streams. As outlined before, the most obvious purpose of this development is

to improve model skill to represent ecosystem dynamics on the NWE Shelf. This is not, however, the only purpose of introducing the bio-optical module into ERSEM, as the spectrally resolved products for underwater light have their own importance, such as for recreational and commercial diving activities and naval operations (*Woodham* [2011]). Furthermore, another crucial aspect of this work is data assimilation, which has been increasingly applied in ecosystem modelling (*Gehlen et al.* [2015]).

Data assimilation is a set of tools and methods that enable us to systematically merge the model forecast with the observational data in order to optimally represent the state of a complex dynamical system. It has been developed mostly in the field of numerical weather forecasting (e.g. *Kalnay* [2003]), but data assimilation has found application in a range of other fields, such as operational oceanography (e.g. *Cummings et al.* [2009]; *Edwards et al.* [2015]). The most typically assimilated data in biogeochemistry are ocean color-derived (satellite) products for surface concentrations of total chlorophyll (*Ishizaka* [1990]; *Carmillet et al.* [2001]; *Natvik and Evensen* [2003]; *Hoteit et al.* [2005]; *Torres et al.* [2006]; *Nerger and Gregg* [2007, 2008]; *Gregg* [2008]; *Fontana et al.* [2010]; *Ciavatta et al.* [2011]; *Ford et al.* [2012]; *Ciavatta et al.* [2016]; *Kalaroni et al.* [2016]; *Ford and Barciela* [2017]; *Pradhan et al.* [2019]), and this has been recently extended to Phytoplankton Functional Type (PFT) chlorophyll (*Ciavatta et al.* [2018]; *Skákala et al.* [2018]; *Ciavatta et al.* [2019]). There is however only a handful of studies that assimilate direct optical radiances, among the few cases there is assimilation of phytoplankton light absorption (*Shulman et al.* [2013]), diffuse light attenuation coefficient (*Ciavatta et al.* [2014]), reflectance data (*Jones et al.* [2016]) and absorption by Coloured Dissolved Organic Carbon (CDOC) (*Gregg and Rousseaux* [2017]). The advantage of using radiances is that they are often directly measured by the satellite and their products have consequently lower errors (e.g *Groom et al.* [2019]). However, whether such products can be assimilated into the model depends on how directly we can relate those radiances to the model state variables. One of the important roles played by the bio-optical module introduced in this work is that it provides the necessary link between spectral radiances and biogeochemistry from the NEMO-FABM-ERSEM model, and therefore increases the capacity of our assimilative systems.

In this study we a) assessed the impact of the new bio-optical module on ERSEM skill in the "free run" (without assimilation) , and b) assimilated satellite PFT chlorophyll (*Brewin et al.* [2017]) and PFT optical absorption (*Brewin et al.* [2019]) products into ERSEM using bio-optical module. The runs assimilating satellite products into ERSEM coupled to the bio-optical module will be compared with two ERSEM free runs (with and without the bio-optical module) and with the currently established assimilation of PFT chlorophyll using the pre-existing ERSEM light scheme (*Ciavatta et al.* [2018]; *Skákala et al.* [2018]). The ERSEM skill in the different simulations will be assessed by looking at how the model represents: a) phytoplankton concentrations, community structure and phytoplankton seasonal cycle (e.g timing and magnitude of the Spring bloom), b) the underwater light field and c) the nutrient cycle. We expect both the bio-optical module and the assimilation to have large impacts on a) and b), with lesser, but still important impacts on c). We also anticipate important changes in the model carbon cycle, but due to lack of data, these can only be evaluated indirectly. In analysing the differences between assimilating PFT chlorophyll and PFT absorption, it is important to note that the products are not independent: PFT absorption is derived from PFT chlorophyll using the model of *Brewin et al.* [2019]. However, the absorption model used in ERSEM (based on *Gregg and Casey* [2009]; *Gregg and Rousseaux* [2016]) is independent from the satellite model of *Brewin et al.* [2019] and so the two may not be entirely consistent. The key role played by data assimilation is that it merges information from multiple sources, and by assimilating PFT absorption into the model we aim to move closer towards the optimal representation of underwater light. However, if there are differences between the ERSEM and satellite absorption models, PFT absorption assimilation might not provide statistically optimal representation of PFT chlorophyll, which in turn should be provided by PFT chlorophyll as-

similation. So even though the PFT absorption data used in this study were derived from PFT chlorophyll, assimilating PFT absorption might have some advantages over assimilating PFT chlorophyll, and vice versa.

2 Model, Data and Methodology

2.1 NEMO-FABM-ERSEM model

2.1.1 The physical component: NEMO

The NEMO ocean physics component is a finite difference, hydrostatic, primitive equation ocean general circulation model (*Madec et al.* [2015]). The NEMO configuration used in this study is to large degree similar to the one used in *Ford et al.* [2017]; *King et al.* [2018]; *Skákala et al.* [2018]: we use the CO6 NEMO version, based on NEMOv3.6, a development of the CO5 configuration explained in detail by *O'Dea et al.* [2017]. The model has 7 km spatial resolution on the Atlantic Meridional Margin (AMM7) domain using a terrain-following $z^* - \sigma$ coordinate system with 51 vertical levels. The lateral boundary conditions for physical variables at the Atlantic boundary were taken from a reanalysis of the GloSea5 Seasonal Forecasting System (*MacLachlan et al.* [2015]); the Baltic boundary values were derived from a reanalysis produced by the Danish Meteorological Institute for the Copernicus Marine Environment Monitoring Service (CMEMS). The model (including biogeochemistry) was initialized from the CMEMS re-analysis produced at the Met Office (product CMEMS-NWS-QUID-004-011, *http://marine.copernicus.eu/services-portfolio/access-to-products/*). The free simulations were performed for a 3 year period (2016-2018) and the more computationally costly assimilative runs for a single year (2016).

The river discharge dataset used by *Ford et al.* [2017]; *Skákala et al.* [2018] has been updated to cover more recent years using data from *Lenhart et al.* [2010]. Unlike *Ford et al.* [2017]; *Skákala et al.* [2018], here the model was forced at the surface by atmospheric fields provided by the high (hourly) temporal and (31 km) spatial resolution realisation (HRES) of Copernicus Climate Change Service (C3S, 2017) ERA-5 reanalysis (*https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5*).

2.1.2 The ecosystem component: ERSEM

ERSEM (*Baretta et al.* [1995]; *Butenschön et al.* [2016]) is a lower trophic level model for marine biogeochemistry, pelagic plankton, and benthic fauna (*Blackford* [1997]). It distinguishes between five chemical components: carbon, chlorophyll, nitrogen, phosphorus and silicon, using variable stoichiometry for the simulated plankton groups (*Geider et al.* [1997]; *Baretta-Bekker et al.* [1997]). The model splits phytoplankton into four functional types largely based on their size (*Baretta et al.* [1995]): picophytoplankton, nanophytoplankton, diatoms and dinoflagellates. Each PFT biomass is represented in terms of chlorophyll, carbon, nitrogen and phosphorus, with diatoms also represented by silicon. ERSEM predators are composed of three zooplankton types (mesozooplankton, microzooplankton and heterotrophic nanoflagellates), with organic material being decomposed by one functional type of heterotrophic bacteria (*Butenschön et al.* [2016]). The ERSEM inorganic component consists of nutrients (nitrate, phosphate, silicate, ammonium and carbon) and dissolved oxygen. The carbonate system is also included in the model (*Artioli et al.* [2012]). We used in this study a well established ERSEM parametrization described in *Butenschön et al.* [2016]. At the Atlantic boundary values for nitrate, phosphate and silicate were taken from World Ocean Atlas (*Garcia et al.* [2013]) and dissolved inorganic carbon from the GLODAP gridded dataset (*Key et al.* [2015]; *Lauvset et al.* [2016]), while plankton and detritus variables were set to constant values.

The pre-existing light scheme in ERSEM is described in *Butenschön et al.* [2016]: the light is taken as diffuse only, and it is forced by the hourly net downwelling short-

wave radiation from the ERA-5 product used to force NEMO as

$$E_{PAR} = q_{PAR} \cdot I_{surf} \cdot \exp\left(\int K_d(z)dz\right), \tag{1}$$

where I_{surf} is the surface downwelling SWR, q_{PAR} is the fraction of PAR and the exponential term describes the broadband light attenuation by the different components in the water. The light attenuation distinguished absorption and backscattering by pure water and the 4 PFTs (based on the model of *Lee et al.* [2005]) as:

$$K_d = (1 + 0.005\theta_{zen}) \cdot a \;+\; 4.18 \cdot (1 - 0.52\exp[-10.8 \cdot a]) \cdot b \tag{2}$$

where θ_{zen} is zenith angle and absorption (a) and backscattering (b) terms are defined as (*Butenschön et al.* [2016]):

$$a = \sum_{i=1}^{4} a_i^* \cdot C_i \;+\; \text{ady} \;+\; a_{\text{sea}} \tag{3}$$

and

$$b = \sum_{i=1}^{4} b_i^* \cdot C_i \;+\; b_{\text{sea}}. \tag{4}$$

The a^*, b^* in Eq.3-4 are specific absorption and backscattering coefficients, i index runs through the 4 PFTs, C is the chlorophyll concentration and $a_{\text{sea}}, b_{\text{sea}}$ are the terms that describe absorption and backscattering by sea water. The ady term captures the absorption by POM, CDOM and sediment which is forced by an external product based on (443 nm wavelength) SeaWIFS data (*Wakelin et al.* [2012]) and derived from the bio-optical model of *Smyth et al.* [2006]. However, the pre-existing ERSEM light scheme does not attempt to provide any genuine representation of underwater light, it merely focuses on estimating the photosynthetic energy flux through the surface of phytoplankton cells.

2.2 Spectrally resolved bio-optical module

The bio-optical module implemented into NEMO-FABM-ERSEM covers both atmosphere and ocean and is externally forced by bulk meteorological properties. The functionality of the atmospheric module matches that of the widely used Ocean-Atmosphere Spectral Irradiance Model (OASIM, *Gregg and Casey* [2009]), but additionally allows for arbitrary spectral resolution and, through FABM, for integration in a large number of oceanographic models. At the ocean surface, the module distinguishes two downwelling radiation streams, diffuse and direct, which are both fully spectrally resolved. These streams are then tracked downwards through the water column as they are being absorbed and scattered by water and ecosystem constituents.

The atmospheric fields driving the module were obtained from the Copernicus Climate Change Service (C3S, 2017) ERA-5 reanalysis ((*https://www.ecmwf.int/*)). The ERA-5 data came with 3-hourly temporal and $1/4°$ spatial resolution and covered the following atmospheric constituents (total aggregates per vertical column): ozone, water vapour, cloud liquid water, cloud cover and the mean sea-level air pressure. These fields were supplemented with data for surface wind speed, air humidity and air temperature, all provided by the NEMO atmospheric (ERA-5) forcing. In addition to these fields we provided the module with aerosol optical thickness at 550 nm from MODerate resolution Imaging Spectroradiometer (MODIS) satellite product with monthly resolution (*https://modis.gsfc.nasa.gov/data/dataprod*).

The underwater irradiance spectra were resolved with 33 wavelengths (between 250 nm and 3700 nm), with each spectral band reduced in the water column through backscattering and absorption by water and PFTs, based on the model of *Gregg and Rousseaux* [2016]:

$$\frac{dE_d(\lambda)}{dz} = -C_d(\lambda)E_d(\lambda) \tag{5}$$

Table 1. The Table shows the inherent optical components distinguished by the model, the data used to calculate light attenuation by those components (e.g. ERSEM output for the PFT chlorophyll-a), together with absorption and scattering spectra and the backscattered fraction used. All spectra are taken from *Gregg and Rousseaux* [2016] (abbreviated as G&R), Fig 3, and the data files accompanying the source code are available at *https://gmao.gsfc.nasa.gov/reanalysis/MERRA-NOBM/software/*.

optical component	source	spectral absorption	spectral scattering	backscattered fraction
water	set as constant	G&R	G&R	0.5
diatoms (chl-a)	ERSEM	G&R diatoms	G&R diatoms	0.002
nanophytoplankton (chl-a)	ERSEM	G&R chlorophytes	G&R chlorophytes	0.0071
picophytoplankton (chl-a)	ERSEM	G&R cyanobacteria	G&R cyanobacteria	0.0032
dinoflaggelates (chl-a)	ERSEM	G&R dinoflaggelates	G&R dinoflaggelates	0.0029
non-living matter	MODIS at 443 nm	G&R CDOC	0	0

and

$$\frac{dE_s(\lambda)}{dz} = -C_s(\lambda)E_s(\lambda) + F_d(\lambda)E_d(\lambda), \tag{6}$$

where E_d is the direct downwelling stream, E_s is the diffuse downwelling stream, the C_d, C_s are light attenuation coefficients and F_d describes forward scattering. In Eq.5-6 we neglected all the upwelling terms from *Gregg and Rousseaux* [2016]. Although backscattering is included in light attenuation, the upwelling stream will be only included in the future version of the spectral module. However, backscattering to total scattering ratio was ≤ 0.007 for all the ERSEM variables (see Tab.1), so the upwelling stream could be reasonably neglected.

The module calculated light attenuation by the water components largely following the model of (*Gregg and Rousseaux* [2016]). The used scheme is summarized in Tab.1. Similarly to the pre-existing ERSEM light scheme the absorption by POM, CDOM and sediment (non-living matter, $a_{nl}(\lambda)$) was forced by an external product extrapolated from wavelength specific (443 nm) data of *Smyth and Artioli* [2010], as:

$$a_{nl}(\lambda) \simeq \exp\{-S(\lambda - 443)\}, \tag{7}$$

with $S = 0.014\ nm^{-1}$, as for Chromophoric Dissolved Organic Carbon (CDOC) in *Gregg and Rousseaux* [2016].

2.3 Data and validation

2.3.1 Assimilated data

We assimilated in this study two satellite products: a PFT chlorophyll (*Brewin et al.* [2017]) and a PFT absorption product (*Brewin et al.* [2019]). Based on a simple, conceptual model (*Brewin et al.* [2010, 2015]), *Brewin et al.* [2017] used the Ocean Colour - Climate Change Initiative (OC-CCI, project of the European Space Agency, Version 3.0, *Sathyendranath et al.* [2016, 2019]) data for total chlorophyll-a to estimate chlorophyll in the 4 ERSEM PFTs (diatoms, dinoflaggelates, nano- and pico-phytoplankton). The model of *Brewin et al.* [2017] splits phytoplankton into a subgroup of smaller species (<20μm, picophytoplankton and nanophytoplankton) and microphytoplankton (diatoms and

dinoflaggelates) as

$$C_{1,2} = C_{1,2}^{m} \cdot \left(1 - \exp\left\{-\frac{D_{1,2}}{C_{1,2}^{m}} \cdot C\right\}\right), \qquad C_{3,4} = C - C_{1,2}, \tag{8}$$

where C is the total chlorophyll concentration, $C_{1,2}$ is the aggregate concentration of pico and nanophytoplankton, $C_{3,4}$ is microphytoplankton, $C_{1,2}^{m}$ is the maximum value of $C_{1,2}$ approached in the asymptotic limit $C \to \infty$, while the $D_{1,2}$ is the fraction $C_{1,2}/C$ in the limit of $C \to 0$. The $C_{1,2}^{m}$ and $D_{1,2}$ parameters are dependent on the Sea Surface Temperature (SST, see *Brewin et al.* [2017]), which was obtained from the satellite data (OISST version from *Reynolds et al.* [2007]). Using analogous model to Eq.8 one can split $C_{1,2}$ further into nanophytoplankton and picophytoplankton concentrations. Furthermore the microphytoplankton concentration can be split into diatoms (C_3) and dinoflaggelates (C_4) using

$$C_3 = \frac{C_{3,4}}{1 - \exp(-\alpha(SST - \beta))}, \qquad C_4 = C_{3,4} - C_3 \tag{9}$$

where α, β are two suitably tuned parameters (*Brewin et al.* [2017]).

The PFT chlorophyll product was already assimilated in *Ciavatta et al.* [2018]; *Skákala et al.* [2018]. It has a daily temporal and 4 km spatial resolution, and comes with bias and uncertainty estimates (in log-space). Both biases and uncertainties were estimated using in situ and satellite data match-ups following the approach from *Jackson et al.* [2017] and fuzzy logic statistics (*Moore et al.* [2009]). It has been demonstrated that the PFT chlorophyll biases and uncertainties depend mostly on the Optical Water Type (OWT, *Brewin et al.* [2017]) with higher OWTs describing the optically complex waters typically found in the coastal and Shelf regions, and in the most dynamical time of the year (i.e. during Spring). As expected, the model has larger uncertainties in the higher OWTs, furthermore the satellite model tends to underestimate chlorophyll concentrations in the lower OWTs and overestimate chlorophyll in the higher OWTs (see *Brewin et al.* [2017]). Similarly to *Ciavatta et al.* [2018]; *Skákala et al.* [2018], we unbiased the satellite data prior to assimilation and calculated the uncertainties of the unbiased data (following the method of *Ciavatta et al.* [2016]).

The PFT absorption product has been derived from the unbiased PFT chlorophyll data using the model of *Brewin et al.* [2019] for the North Atlantic. Based on specific absorption coefficients $a_i^*(\lambda)$ fitted from in situ measurements, the model of *Brewin et al.* [2019] derives PFT absorption ($a_i(\lambda)$) for 12 characteristic wavelengths (λ) as

$$a_i(\lambda) = a_i^*(\lambda) \cdot C_i \tag{10}$$

where the i index labels the specific PFT. 6 out of these 12 wavelengths were selected for the satellite product: 412 nm, 443 nm, 490 nm, 510 nm, 555 nm and 665 nm. Since the PFT absorption is derived from PFT chlorophyll, it has the same spatial and temporal resolution as PFT chlorophyll. Similarly to PFT chlorophyll, PFT absorption also contains information about uncertainties, which have two sources: the original uncertainty of the PFT chlorophyll and additional uncertainty associated with the specific absorption coefficients (*Brewin et al.* [2019]). PFT absorption therefore has larger uncertainty than PFT chlorophyll.

2.3.2 Validation data for the oceanic variables

We used both shelf-wide and location-specific in situ measured data to assess the model skill. For the shelf-wide data we used three data-sets: a) the Ecosystem Data Online Warehouse of the International Council for the Exploration of the Sea (ICES, *https://www.ices.dk/marine-data/*) that contains measurements of three nutrients of specific interest: nitrate, phosphate, silicate, and also data for total chlorophyll. The ICES data-set for year 2016 contained a large number of measurements (for each variable around 5000)

at a range of depths (0 - 30m) at different locations on the NWE Shelf (southern North Sea, eastern UK coastline, Irish coastline). b) We also used the 1960-2014 monthly climatologies for total chlorophyll, nitrate, phosphate and silicate collected in the North Sea Biogeochemical Climatology (NSBC) project (*Hinrichs et al.* [2017]). The NSBC data-set is gridded with $1/4°$ horizontal resolution and 16 vertical layers, covering most of the North Sea domain. Although it is methodologically difficult to validate model skill with climatologies, the NSBC data-set is a valuable source of information due to its extensive spatial coverage. c) To validate surface light attenuation coefficient (Kd) at 490 nm wavelength we used the merged OC-CCI satellite data-set of the European Space Agency (ESA), product version 4.1 (*http://www.esa-oceancolour-cci.org/*).

The location-specific data were obtained for the long-term monitoring station L4 in the English Channel (*https://www.westernchannelobservatory.org.uk*). To support some of the conclusions we looked at two types of data: weekly climatology derived from 1994-2015 time series for total phytoplankton carbon (*Widdicombe et al.* [2010]) and the underwater PAR for year 2016. The phytoplankton carbon time-series was obtained from a measurement location at 10m depth, while the PAR data were measured across the whole water column (0 - 50m). At each day, the PAR data were collected at a specific time in the morning. Since PAR is highly variable throughout the day (diurnal cycle, cloudiness), the observations cannot be directly compared to the model daily average outputs. However, PAR attenuation in the water column is much more stable on the daily time scale, as it depends mainly on the biogeochemical state of the water, which typically evolves on supra-daily scales (with some notable exceptions, such as migration of dinoflaggelates). Therefore, we compared model and observations as ratios between the PAR values at a range of depths and the PAR value at 2.4 m. PAR data at depth shallower than 2.4 m were excluded due to insufficient number of observations.

2.3.3 Skill metrics

Validating a model with in situ data is rarely trivial due to spatio-temporal differences in model and data resolution (e.g *Schutgens et al.* [2017]). In general it is expected that bias (difference in model and data mean values) is reasonably unaffected by the different resolutions, but Root Mean Square Difference

$$RMSD = \sqrt{\overline{(\text{model} - \text{observations})^2}} \tag{11}$$

can be substantially increased by the in situ small-scale variability. The impact of small scale variability can be reduced by suitably binning the observations. This paper focuses on two metrics: bias and Bias-Corrected RMSD (BC RMSD), which is the same as RMSD from Eq.11, but with bias subtracted from the model. To correct the RMSD for small scale in situ variability we decided to bin the in situ data along two dimensions with the largest data variability: the temporal and the vertical dimension. We calculated (bias corrected) RMSD along both temporal (with monthly bins) and vertical axes, and the total (bias corrected) RMSD presented in the paper is a simple average between those. However, this approach was applied only to the ICES data-set, as the NSBC monthly climatology has larger spatio-temporal resolution than the model. Comparing the model with the gridded NSBC and satellite data is more straightforward, as all it needs is re-gridding the finer resolution data-set on the coarser resolution scale.

2.3.4 Validation of the atmospheric part of the bio-optical module

The bio-optical module was initially validated in the 1D 2013-2016 simulation by the atmospheric data provided by the Western Channel Observatory at the L4 station in the English Channel (some selected results for the atmospheric L4 fields are shown in Tab.2). For the whole NWE Shelf, the module is validated in Fig.1 and Fig.2. These two Figures compare the total downwelling Short-Wave Radiation (SWR) above the ocean

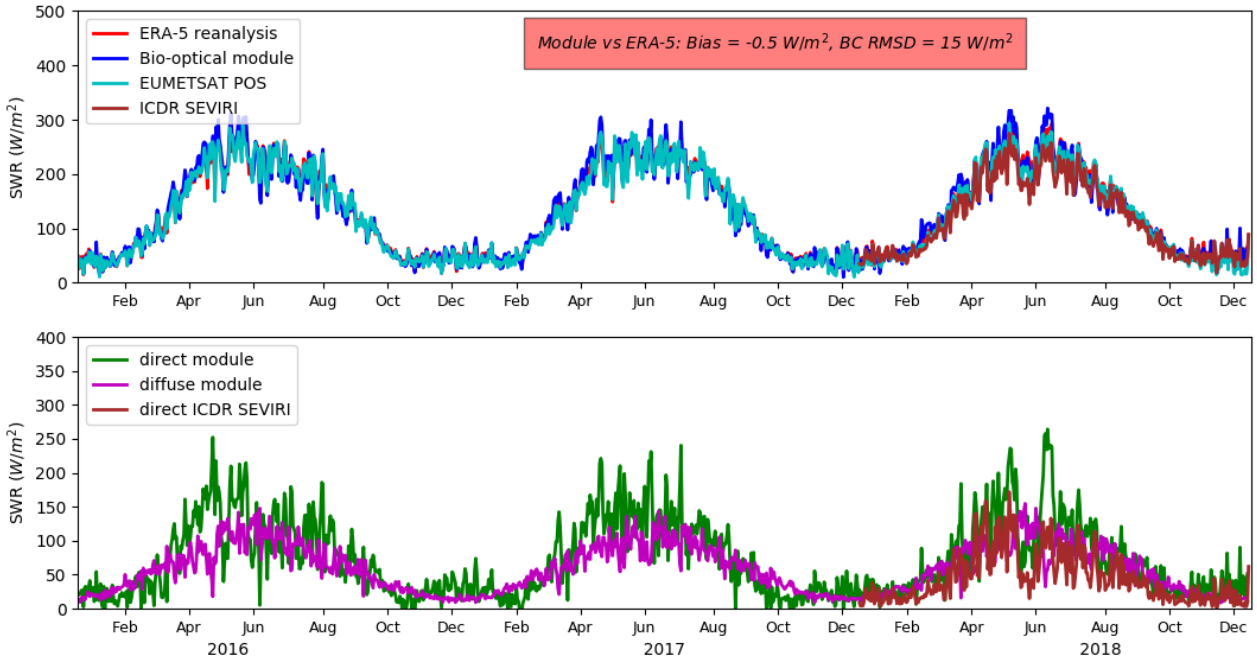


Figure 1. The upper panel shows 2016-2018 time series for the incoming total flux of Short Wave Radiation (SWR) at the ocean surface (before albedo) averaged through the NWE Shelf. We compare the output of the bio-optical module with the ERA-5 reanalysis, which is used to force the pre-existing ERSEM light scheme. We also include the EUMETSAT Polar Orbiting Satellite (POS) SWR and for 2018 the Interim Climate Data Recors (ICDR) SEVIRI data. The bottom panel shows the total surface SWR from the bio-optical module split into direct and diffuse radiation. The direct radiation is for 2018 compared with the ICDR SEVIRI direct downwelling radiation satellite data.

Table 2. The skill score (bias and Bias-Corrected RMSD, BC RMSD) for module and ERA-5 when compared to multiple satellite products: EUMETSAT POS (SWR), ICDR SEVRI data (SWR), a MODIS-Aqua data (PAR) and ICDR SEVRI data for direct downwelling SWR (DIR SWR). The bottom three rows compare module within 1D 2013-2016 simulation with the L4 observations for SWR and diffuse SWR (DIFF SWR). The relative (%) values in brackets are the skill score compared to the median observed value. The POS SWR and MODIS PAR data are compared throughout the NWE Shelf for the full 2016-2018 period, whilst the ICDR SEVRI data were available only for the year 2018.

data	Module bias	Module BC RMSD	ERA-5 bias	ERA-5 BC RMSD
POS SWR in W/m^2	-0.9 (-0.7%)	22 (16.2%)	2.4 (1.8%)	16.8 (12.4%)
SEVIRI SWR in W/m^2	12.7 (9.7%)	22.6 (17.3%)	15.8 (12.1%)	19.6 (14.9%)
MODIS PAR in bar/(m^2day)	-0.1 (-0.7%)	3.7 (20.8%)	–	–
SEVIRI DIR SWR in W/m^2	12.1 (39%)	17 (54.8%)	–	–
L4 SWR in W/m^2	-9.2 (-6%)	18.7 (12%)	-1.3 (-1%)	15.7 (10%)
L4 DIFF SWR in W/m^2	3.2 (5%)	13.2 (20%)	–	–

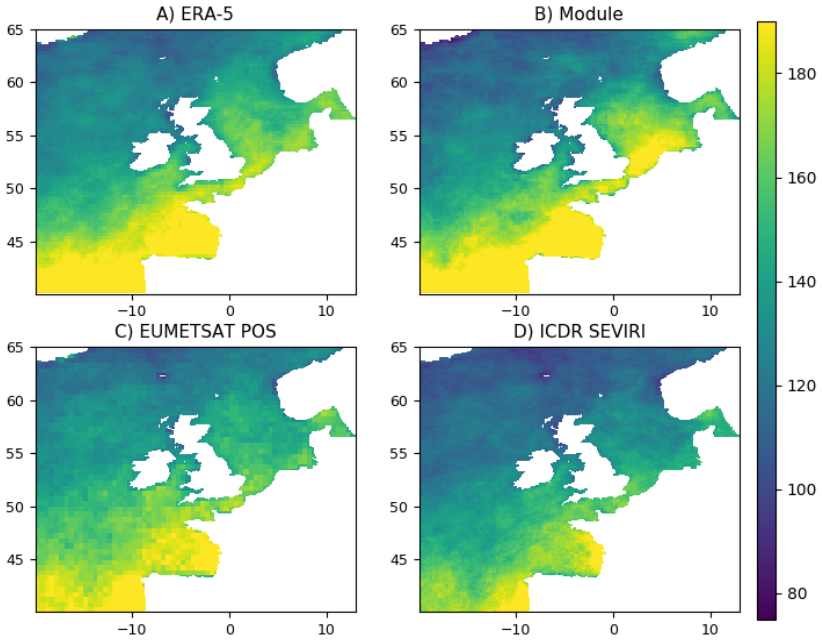


Figure 2. The spatial distribution of 2018 median total downwelling Short Wave Radiation (SWR) flux (at the ocean surface and before albedo, W/m^2) compared between the bio-optical module (A), the ERA-5 re-analysis (B) used to force the pre-existing ERSEM light scheme, and two satellite products, the EUMETSAT POS (C) and ICDR SEVIRI data (D).

surface between the module output, the ERA-5 product, which is used to force the pre-existing ERSEM light scheme, the Interim Climate Data Record (ICDR) SEVIRI sensor product v400/v410 (*https://wui.cmsaf.eu/*) and the EUropean organisation for the exploitation of MeTeorological SATellites (EUMETSAT) incoming daily SWR product from Polar Orbiting Satellites (POS, version 1.9.5, *http://mpimet.mpg.de/cdi*). It is shown (Fig.1 and Fig.2) that the module is reasonably consistent in terms of temporal and spatial patterns with the ERA-5 data provided by the CMEMS reanalysis. The relative bias between the bio-optical module and the ERA-5 product (module minus ERA-5) is very small (-0.5 W/m^2, < 1%), with Bias-Corrected Root Mean Square Difference (BC RMSD, Eq.11) around 10% from the median ERA-5 SWR value.

The Tab.2 shows two skill metrics (bias and BC RMSD, Eq.11) evaluating performance of both module and ERA-5 SWR when compared to the ICDR SEVIRI and EUMETSAT POS satellite data. The two satellite SWR products are not entirely consistent: the EUMETSAT POS SWR matches nicely with both ERA-5 and the module, but all are on average about 10% larger than ICDR SEVIRI data. In both cases the module slightly outperforms ERA-5 in bias, but slightly underperforms in BC RMSD. The latter is probably due to larger spatial variability in module SWR when compared to both the ERA5 and the two satellite products (Fig.2). Tab.2 further compares the module PAR with MODIS-Aqua Level 3 satellite PAR product (*https://oceancolor.gsfc.nasa.gov/l3/*) showing very similar skill score to the comparison with EUMETSAT POS in SWR (negligible bias of -0.7%). The last validation data-set shown in Tab.2 was ICDR SEVIRI product for direct downwelling SWR. Not surprisingly, module skill in representing direct downwelling SWR is consistent with module skill in representing SWR when validated with the same ICDR SEVIRI data. The module relative bias was larger for direct SWR than for total SWR, suggesting that the overestimated direct SWR is the main reason why module overestimated total SWR relative to ICDR SEVIRI data.

2.4 The Data Assimilation (DA) system

We used the data assimilation set-up already described in *Skákala et al.* [2018], which is based on NEMOVAR (*Mogensen et al.* [2009, 2012]; *Waters et al.* [2015]), a 3D-VAR variational DA system used for operational ocean DA at the UK Met Office. The 3D-VAR version applied in this study uses the First Guess at Appropriate Time (FGAT) approach and minimizes the cost function using the conjugate gradient method (*Mogensen et al.* [2012]). DA of PFT chlorophyll into NEMO-FABM-ERSEM using NEMOVAR has been implemented at the Met Office for use in reanalysis and in the future it will be considered for operational forecasting (see *Skákala et al.* [2018]). The scheme starts with univariate assimilation of four separate PFTs (diatoms, nanophytoplankton, picophytoplankton, dinoflagellates) surface chlorophyll concentrations. The analysis is performed in log-space, taking account of the typical log-normal distribution of chlorophyll concentrations in the ocean (*Campbell* [1995]). The surface PFT chlorophyll increments are propagated with constant values vertically within the mixed layer. After calculating the increments for PFT chlorophyll, one may use a balancing module to update some additional ERSEM state variables. In the case of PFT chlorophyll assimilation, it is essential to preserve the background stoichiometric ratios between the PFT components (chlorophyll, carbon, nitrogen, phosphorus, silicon), as those ratios reflect on the physiological adaptation of the PFT cells to the environment. The balancing scheme therefore updates the PFT chemical components (other than chlorophyll) as

$$\Delta X_i = (X_i/C_i) \cdot \Delta C_i, \tag{12}$$

where i labels the 4 PFTs, X_i is a given chemical component of a PFT, C_i is the PFT chlorophyll, and $\Delta C_i, \Delta X_i$ are increments of C_i, X_i respectively.

The PFT absorption satellite data (*Brewin et al.* [2019]) were available for 6 wavelengths (412 nm, 443 nm, 490 nm, 510 nm, 555 nm, 665 nm) and those did not exactly

match with the 33 wavelengths used by the bio-optical module. In order to match the bio-optical module with the satellite product, the PFT absorption values corresponding to the 6 satellite wavelengths were linearly interpolated from the module outputs prior to assimilation. The univariate assimilation was applied to calculate increments for the 24 (4 PFTs times 6 wavelengths) surface radiances. Since PFT absorption is only a diagnostic variable, PFT absorption increments will have no impact on the model unless they are translated into increments of some model state variables. The model state variables with straightforward relationship to PFT absorption are the PFT chemical components (chlorophyll, carbon, nitrogen, phosphorus and silicon). It is natural to first update the PFT chlorophyll by transforming PFT absorption increments into PFT chlorophyll increments through the PFT per wavelength specific absorption coefficients (see Eq.10). In principle different wavelengths could produce different PFT chlorophyll increments, so the unique PFT chlorophyll increment was obtained as an average through the PFT chlorophyll increments calculated from the 6 wavelengths. After the PFT chlorophyll increments were calculated from the absorption increments, the same balancing scheme as in PFT chlorophyll assimilation provided the increments for the remaining phytoplankton components.

For both PFT chlorophyll (*Brewin et al.* [2017]) and PFT absorption (*Brewin et al.* [2019]) the per-pixel observation errors were provided with the satellite products. The background errors were estimated by a) binning the data monthly and into four characteristic regions based on bathymetry: 1) region with < 20 m depth, 2) $20 - 50$ m depth, 3) $50 - 200$ m depth, 4) > 200 m depth, and b) by assuming that background and observation errors can be treated inside each bin as independent. Using a) and b) we estimated the background errors by subtracting observation errors from the unbiased differences between model and observations within each bin. A similar scheme is used typically for model diagnostics (e.g *Andersson* [2003]; *Desroziers et al.* [2005]), however due to limited computational resources we had to estimate the background errors from the free run instead of reanalysis. Such simplified scheme has been found sufficient in this application. In fact, in a previous study (*Skákala et al.* [2018]) where PFT chlorophyll was assimilated into NEMO-FABM-ERSEM with NEMOVAR, we found that the precise formulation of error covariances had relatively small impact on the reanalysis, when compared to some other NEMOVAR system features, especially the formulation of the balancing scheme. Overall the background-to-observational error ratios varied between different months and the four bathymetric regions, but on average the background errors were 2 to 3 times larger than the observational errors. This is consistent with what has been found in previous studies (e.g *Ford and Barciela* [2017]).

3 Results and Discussion

3.1 Biogeochemistry: free runs

Fig.3 shows the impact of the bio-optical module on the NWE Shelf phytoplankton and nutrient seasonal cycle. The PFT and total chlorophyll time series (Fig.3) follow the pattern already found in the literature (*Skákala et al.* [2018]): the model has much stronger seasonality than the satellite data and NSBC in situ climatology, with late Autumn - early Spring model chlorophyll concentrations far beneath the observed values. Both free model runs display late blooms (delayed by $\sim$ 1 month) compared to both satellite and in-situ NSBC climatology (Fig.3). The bio-optical module has only moderate impact on the chlorophyll time series, by further increasing both chlorophyll concentrations and model seasonality, while at the same time increasing surface nutrient concentrations (Fig.3). The impact of the bio-optical module on surface nutrient concentrations can be better understood through Fig.4, which shows the values of total phytoplankton chlorophyll averaged in 3D space (i.e. across the NWE Shelf) and time (year 2016), as well as PFT chlorophyll, total phytoplankton and total zooplankton carbon, and nutrients (nitrate and phosphate). Fig.4 shows that even though the bio-optical module moderately increases chlorophyll, it results in lower concentrations of phytoplankton and zooplankton (carbon)

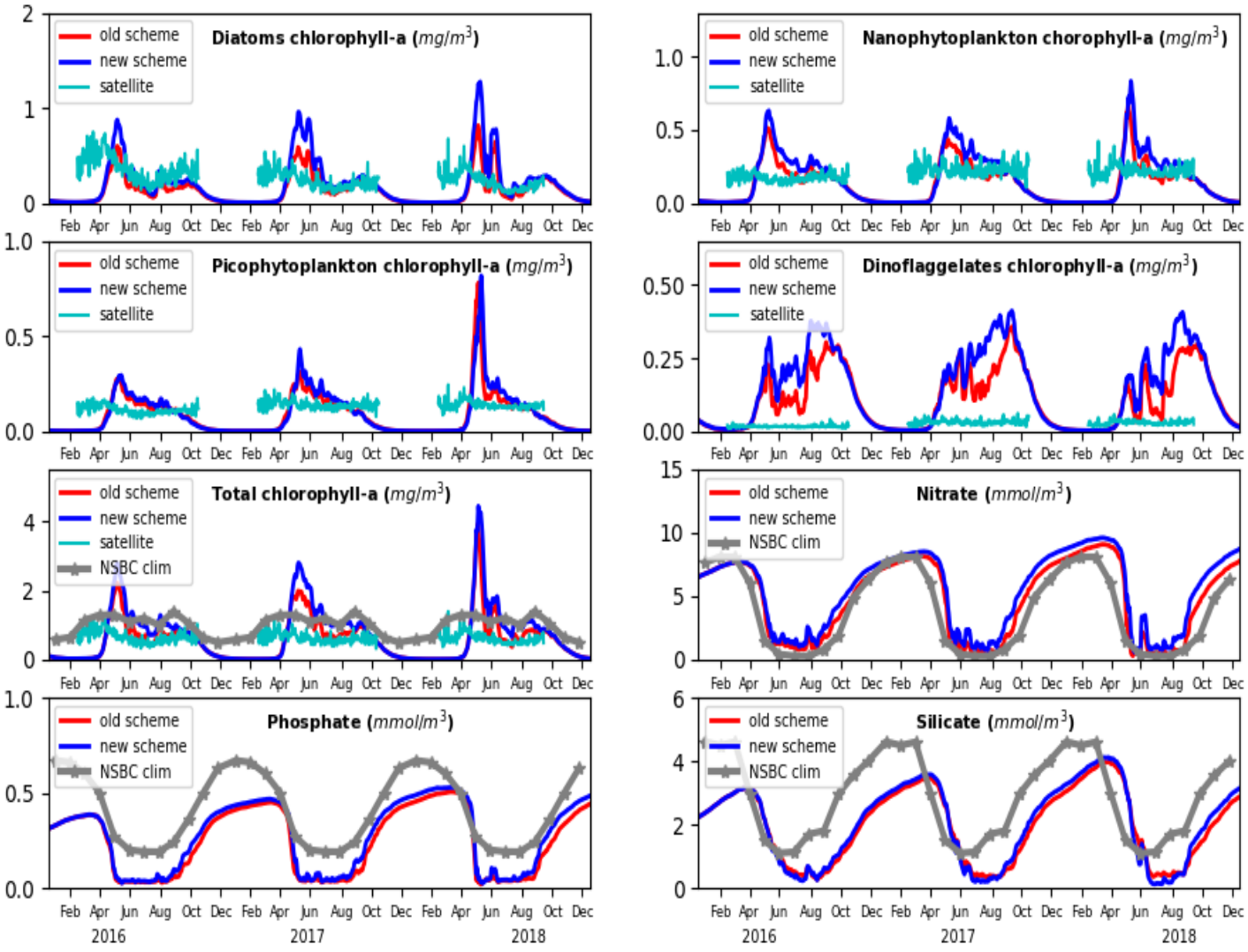


Figure 3. The 2016-2018 time-series for spatially averaged surface concentrations on the NWE Shelf for phytoplankton chlorophyll and nutrients. The Figure compares a) the free run using the pre-existing light scheme ("old scheme"), b) the free run using the new bio-optical module ("new scheme") with either Ocean Color (OC) satellite data (PFT and total chlorophyll), or NSBC climatological data-set (total chlorophyll and nutrients). Since the satellite data between November-February are sparse and located entirely in the south of the domain, the corresponding satellite chlorophyll time-series have been removed.

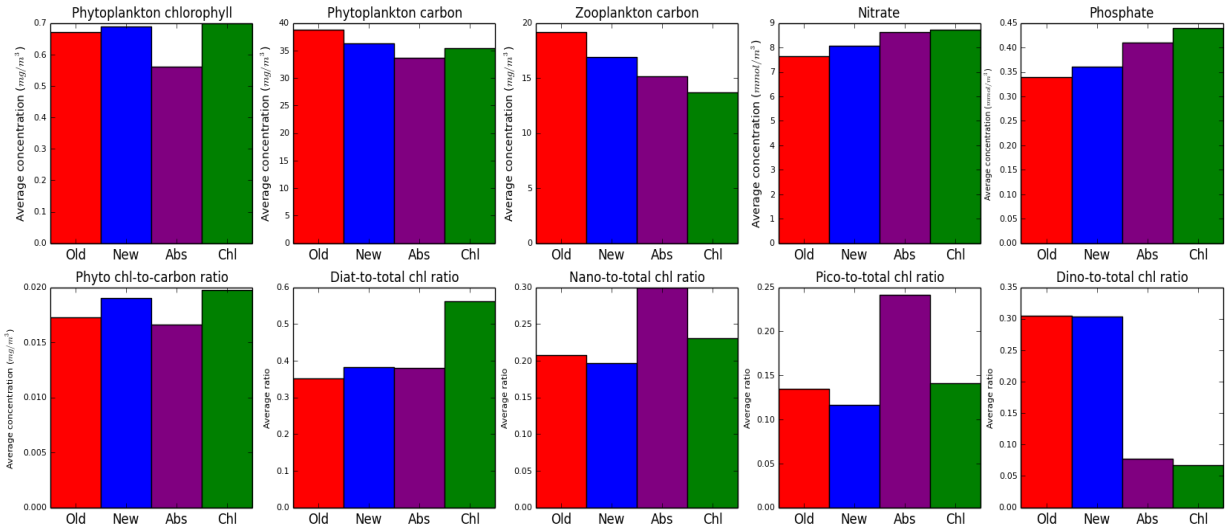


Figure 4. The ecosystem indicators averaged through the year 2016 and the whole Shelf domain (including vertical dimension). The chart compares the values for the different simulations: a) the free run using the pre-existing light scheme ("Old"), b) the free run using the bio-optical module ("New"), c) the run assimilating PFT absorption and using the new bio-optical module ("Abs"), d) the run assimilating PFT chlorophyll-a and using the new bio-optical module ("Chl"). The panels show total phytoplankton and total zooplankton carbon biomass, nutrients (nitrate and phosphate), total chlorophyll-to-carbon ratio and the PFT community structure (PFT-to-total chlorophyll ratio). The PFT abbreviations are: "Diat": Diatoms, "Nano": Nanophytoplankton, "Pico": Picophytoplankton and "Dino": Dinoflagellates.

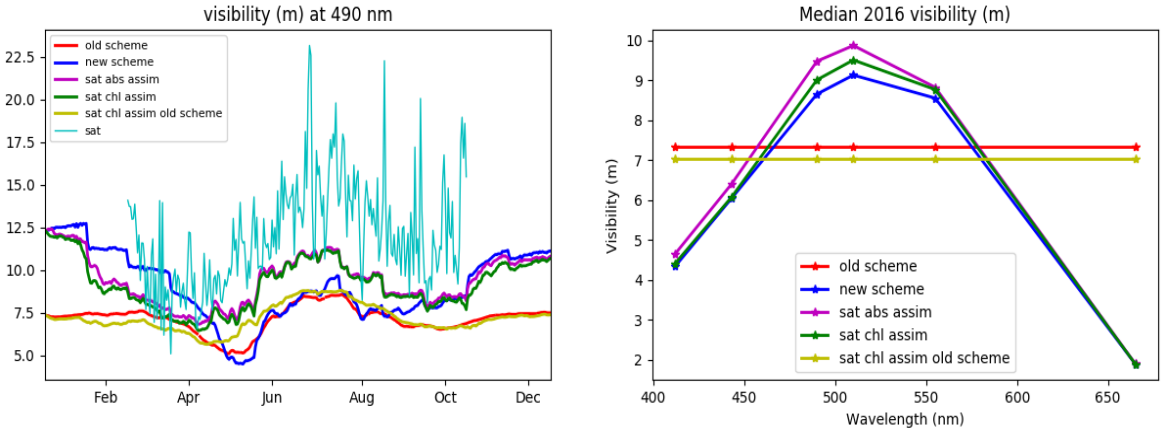


Figure 5. The left hand panel shows visibility (in meters, defined as 1/Kd) for 490 nm wavelength at the ocean surface. As previously, the panel shows 2016 time series of the spatially averaged value across the NWE Shelf. The panel compares a) free run using pre-existing light scheme ("old scheme"), b) free run using bio-optical module ("new scheme"), c) the run assimilating PFT satellite absorption and using bio-optical module ("sat abs assim"), d) the run assimilating PFT chlorophyll and using bio-optical module ("sat chl assim") and e) the run assimilating PFT chlorophyll and using pre-existing light scheme ("sat chl assim old scheme"). The runs are compared with the OC satellite data (as previously the November-February satellite time-series were removed due to data sparsity). The right hand panel shows 2016 Shelf median ocean surface visibility for 6 outputted wavelengths. Since the visibility of the pre-existing light scheme is taken as broad-band, the most consistent way of comparing it to the spectrally resolved visibility of the bio-optical module is to represent it across the spectral band with a constant value.

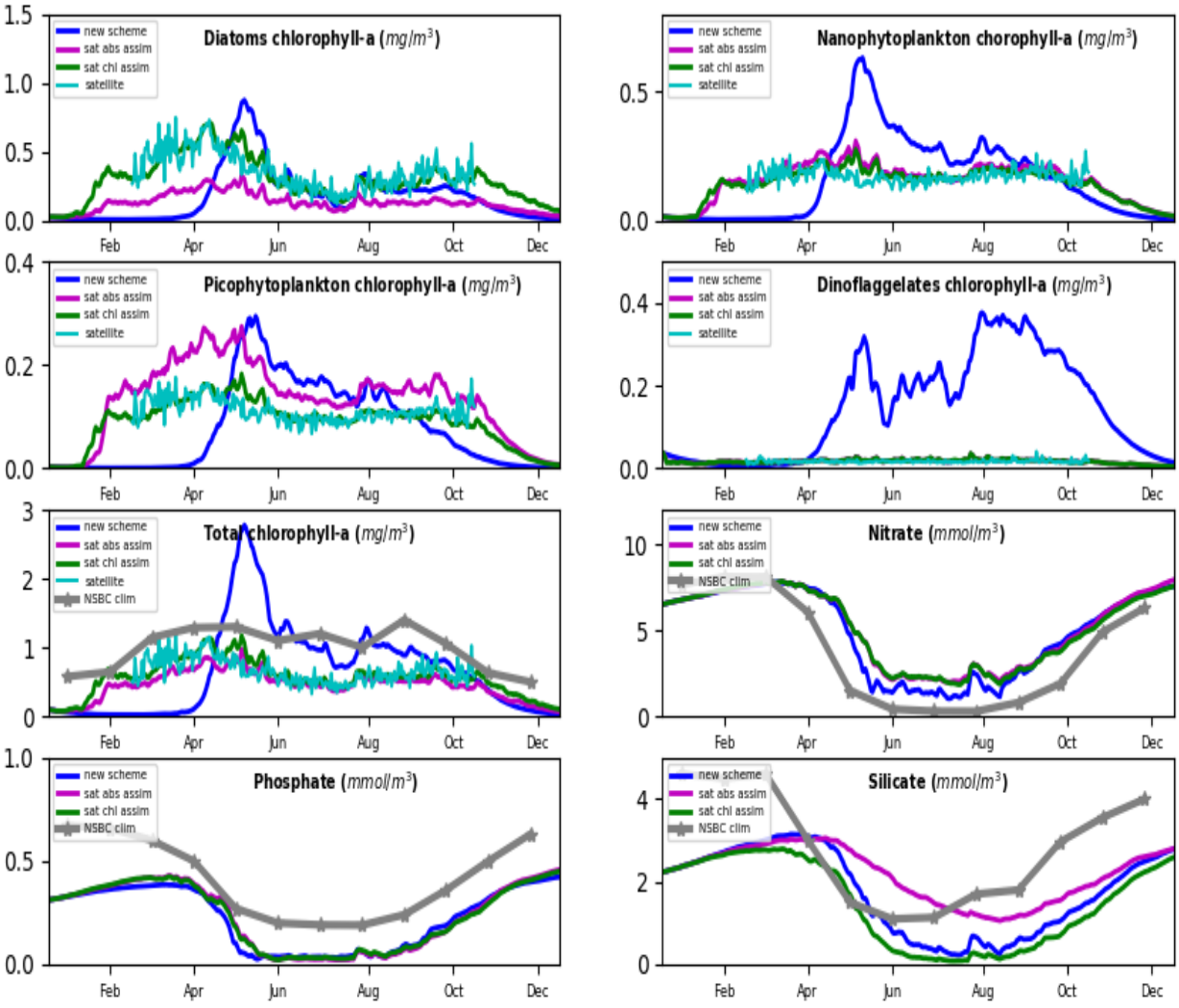
biomass, leading to an overall increase in nutrients. This is most likely due to lower primary productivity caused by reduced photosynthetic radiation in the water column, which was found to be 10-20% lower (depending on the season) in the bio-optical module than in the pre-existing ERSEM light scheme. Since there is negligible bias in the incoming solar irradiance (Fig.1), we expect that the bio-optical module reduces underwater PAR due to increased light extinction inside the watercolumn. Fig.5 suggests that the module reduces underwater PAR dominantly in two wavebands: the $\sim$ 400 - 470 nm waveband, which is mostly absorbed by detritus and particulate matter, and the 570 - 700 nm waveband, which is mostly absorbed by the sea water (e.g *Gregg and Casey* [2009]; *Gregg and Rousseaux* [2016]).

The PFT chlorophyll-to-carbon ratio is a good indicator of the environmental (nutrients, irradiance, temperature) impact on PFT cell physiology (e.g *De Mora et al.* [2013]), with darker environments producing larger chlorophyll values relative to carbon (*Finenko et al.* [2003]). We indeed observed (not shown here) that the reduced PAR in the bio-optical module lead to an overall increase of the PFT chlorophyll-to-carbon ratios. Since the module did not substantially change the PFT community structure (Fig.4), the larger PFT chlorophyll-to-carbon ratios explain the increase in the total phytoplankton chlorophyll-to-carbon ratio from the Fig.4.

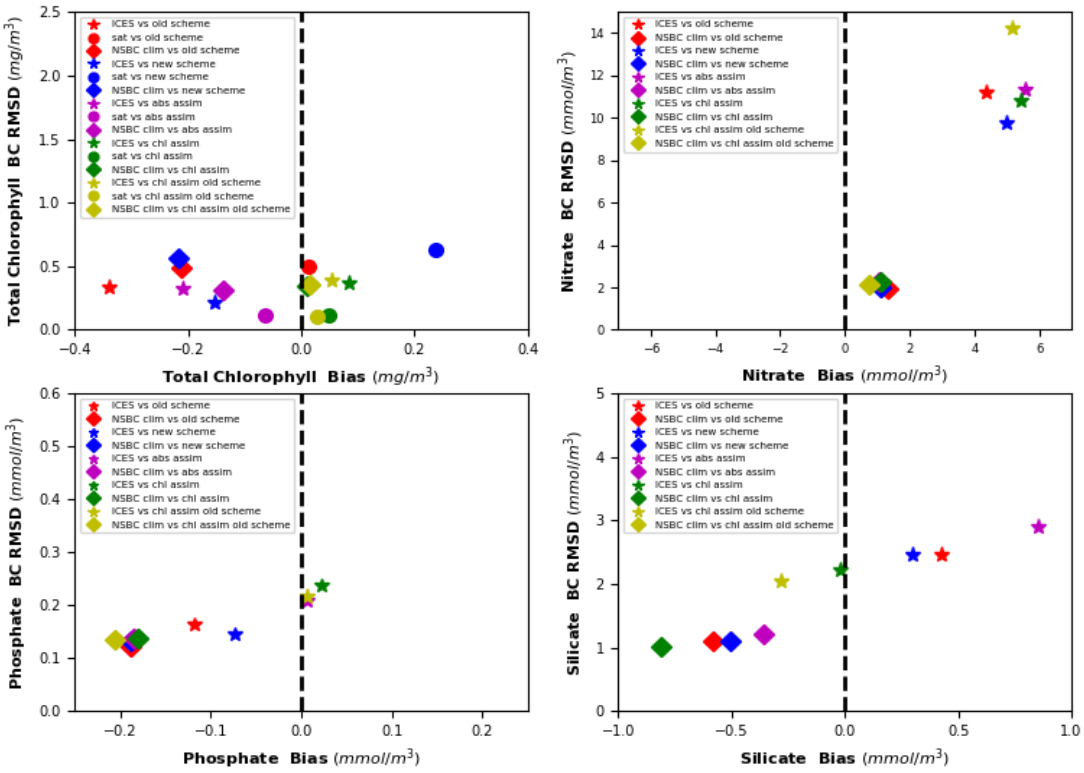
3.2 Biogeochemistry: assimilative runs

Fig.6 shows that the assimilation of satellite PFT chlorophyll and satellite PFT absorption into the bio-optical module has large impact on PFT community structure and substantially improves the seasonal patterns of the phytoplankton growth. In particular: a) assimilation moderates the extremity of the model Spring bloom and b) it moves the Spring bloom by around $\sim$ 1 month towards the start of the year. This is consistent not only with the assimilated satellite data, but also with the NSBC climatology, as well as with the seasonality observed in the ICES data (not shown here, but for overall skill score see Fig.7). However, the light module has only little impact on this improvement in model skill. For example assimilating PFT chlorophyll using the pre-existing light scheme (*Skákala et al.* [2018]) carries similar skill (Fig.7) than the two assimilative runs shown in Fig.6. The changed phytoplankton dynamics in the assimilative runs produces similar annual mean chlorophyll than the free runs (Fig.4). The difference in total chlorophyll-a is largest between the two assimilative runs (Fig.4), which is due to differences in the specific absorption coefficients used in the satellite algorithm and in the bio-optical module (Fig.8). As shown in Fig.8, the largest differences in the specific absorption coefficients are for diatoms and picophytoplankton. These differences are responsible for the distinct PFT community structure between the two assimilative runs (Fig.4 and Fig.6). In particular the relatively lower absorption of picophytoplankton implied by the bio-optical module (Fig.8) produces for PFT absorption assimilation higher picophytoplankton concentrations than the one produced in the PFT chlorophyll assimilative run (Fig.6). And vice versa, the relatively higher absorption by diatoms (Fig.8) in the bio-optical module increases diatoms concentrations in the PFT absorption assimilative run (Fig.6), when compared to the PFT chlorophyll assimilation. Since diatoms are silicate users, the changed diatom concentrations between the two assimilative runs have substantial impact on silicate concentrations (Fig.6). Overall, nutrients have the largest concentrations in the assimilative runs, which is explained by the lower plankton biomass in the assimilative runs relative to the free runs (Fig.4).

There are only few observation data for phytoplankton other than chlorophyll, so it is hard to determine whether the changed plankton biomass improves, or degrades the model skill. Some indication can be obtained from the data at the specific L4 location in the English Channel, where longer time series exist for total phytoplankton carbon. The L4 phytoplankton carbon time series were available for the period 1992-2015, just short of the simulation year of 2016, but it is still possible to compare (Fig.9) the L4 phyto-



549 **Figure 6.** The 2016 time-series for spatially averaged surface concentrations on the NWE Shelf for phyto-
550 plankton chlorophyll and nutrients. The Figure compares a) the free run using the bio-optical module ("new
551 scheme"), b) the run assimilating PFT absorption and using the new bio-optical module ("sat abs assim"), c)
552 the run assimilating PFT chlorophyll and using the new bio-optical module ("sat chl assim"), with either OC
553 satellite data (PFT and total chlorophyll), or NSBC climatological data-set (total chlorophyll and nutrients).
554 The satellite time series for November-February have been removed due to data sparsity.



555 **Figure 7.** The skill score (bias vs BC RMSD) in representing total chlorophyll and nutrients (nitrate, phos-
556 phate, silicate) for the different simulations: a) free run using pre-existing light scheme (red), b) free run using
557 bio-optical module (blue), c) the run assimilating PFT satellite absorption and using bio-optical module (pur-
558 ple), d) the run assimilating PFT chlorophyll and using bio-optical module (green) and e) the run assimilating
559 PFT chlorophyll and using pre-existing light scheme (yellow). The different markers represent comparison
560 with different data: ICES data-set (star), OC satellite product (circle) and NSBC climatology (diamond).

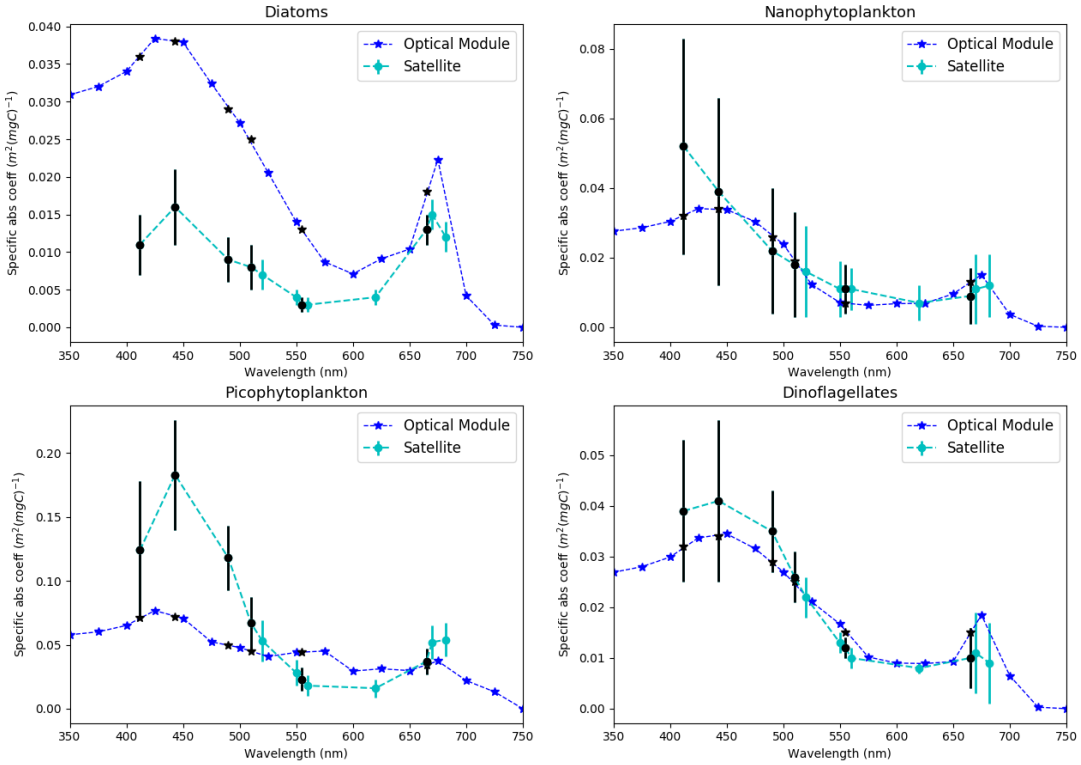


Figure 8. Comparison of PFT specific absorption coefficients a_i^* from the satellite model of *Brewin et al.* [2019] with the PFT specific absorption coefficients used in the bio-optical module (i.e *Gregg and Rousseaux* [2017]). The black markers show a) the selected six wavelengths used in the satellite product and b) the corresponding six interpolated specific absorption coefficients from the module.

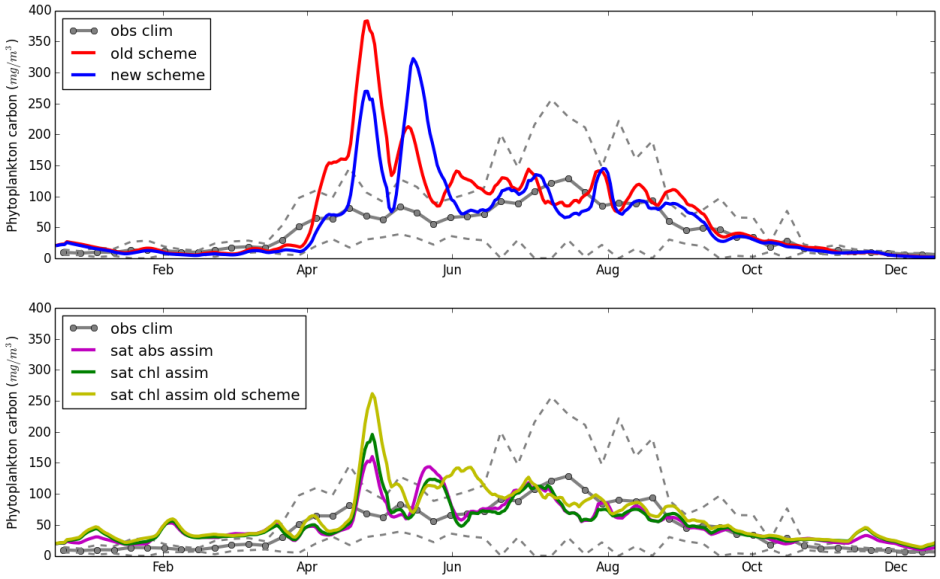
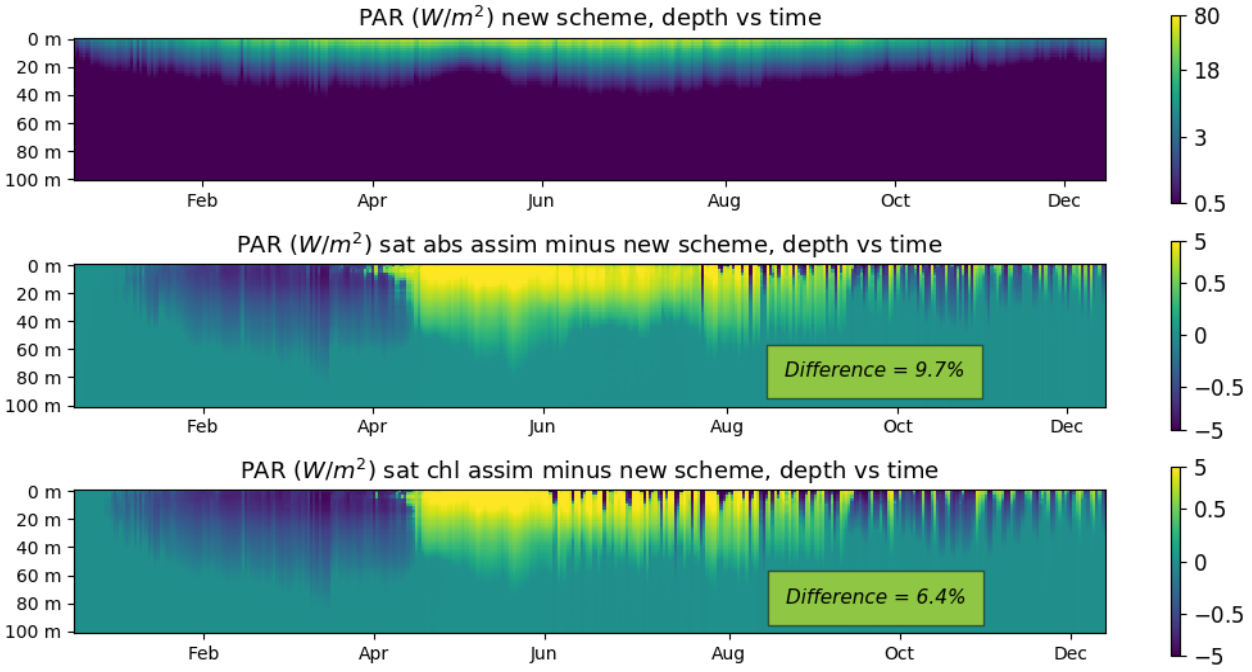


Figure 9. The Figure compares total phytoplankton carbon at the L4 location at 10 m depth. Since the L4 observations were available only for 1992-2015 period, the Figure shows 1992-2015 L4 weekly climatology ("obs clim"). The dashed lines show the interval corresponding to the inter-annual variability of the L4 data (for each week the dashed lines represent ± standard deviation around the mean). The L4 phytoplankton carbon data showed no trend, so the observational climatology can be reasonably compared to the model 2016 time-series. To avoid the Figure being too crowded we split the time-series into two panels, with the upper panel comparing the two free runs: the run using pre-existing light module ("old scheme") and the run using bio-optical module ("new scheme"), with the L4 data. The bottom panel compares the assimilative runs: PFT absorption assimilation using the bio-optical module ("sat abs assim"), the PFT chlorophyll assimilation using the bio-optical module ("sat chl assim") and the PFT chlorophyll assimilation using the pre-existing light scheme ("sat chl assim old scheme"), with the L4 data.



620 **Figure 10.** Comparison of PAR (400 - 700 nm) in the water column. The panels show 2016 time series
621 (x-axis) of median values for the NWE Shelf at each depth (y-axis, 0-100 m range). The upper panel shows
622 PAR for the free run using the bio-optical module, the middle panel shows the difference between the run
623 assimilating PFT absorption into bio-optical module and the free run using the bio-optical module, and the
624 bottom panel shows the same as the middle panel, but for PFT chlorophyll assimilation. Although the daily
625 differences are large due to changed phytoplankton seasonal cycle, the year-averaged differences are smaller
626 and are shown (in %) within the yellow boxes that appear on the panels.

609 plankton carbon climatology derived from 1992-2015 data to the (year 2016) L4 values
610 extracted for each model run. Fig.9 suggests that the model that uses the pre-existing light
611 scheme overestimates primary productivity, with some indications of a small improvement
612 in model skill carried by the bio-optical module (Fig.9). A much more substantial im-
613 provement in model skill is then carried by the assimilation of PFT absorption, or PFT
614 chlorophyll using the bio-optical module (Fig.9), which both also outperform PFT chloro-
615 phyll assimilation using the pre-existing light scheme (Fig.9). Although the results pre-
616 sented in Fig.9 are location-specific, the trend in carbon concentrations between the dif-
617 ferent simulations observed in Fig.9 copies the trend from Fig.4. Fig.9 might therefore
618 indicate that the bio-optical module has a positive impact on how the model represents the
619 carbon cycle.

627 ### 3.3 The underwater light

634 Fig.10 compares the underwater PAR between the free run using the bio-optical
635 module and the two corresponding assimilative runs. It is shown (Fig.10) that the PAR
636 in the watercolumn has been to some degree increased by PFT absorption assimilation (~

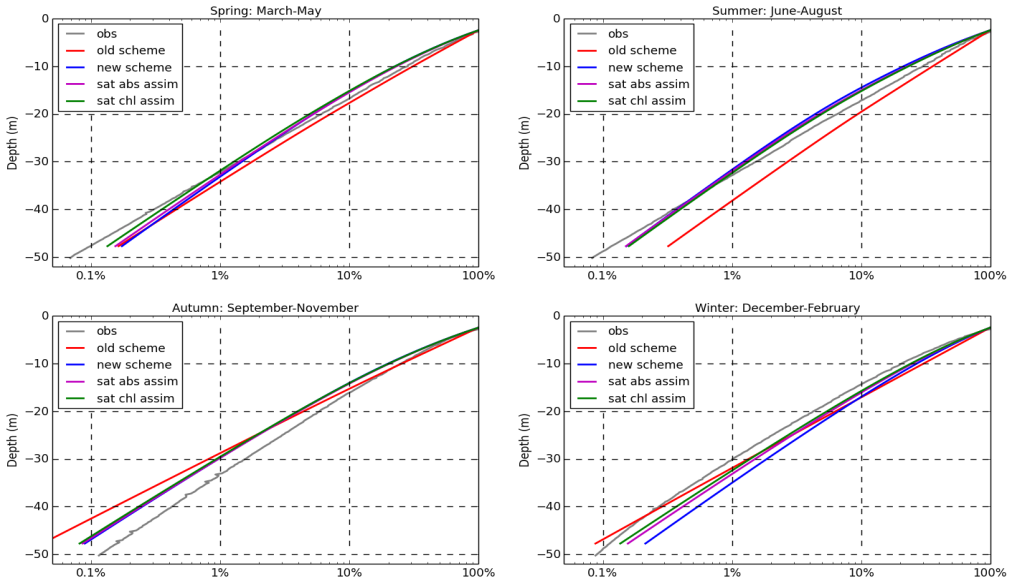


Figure 11. The Figure shows for the L4 location how PAR reduces in the water column relative to its value at the 2.4 m depth (observation data above 2.4 m were too sparse to be used). The Figure compares observed data with four model runs: free run using pre-existing light scheme ("old scheme"), free run using bio-optical module ("new scheme"), PFT absorption assimilative run using bio-optical module ("sat abs assim") and PFT chlorophyll assimilative run using bio-optical module ("sat chl assim"). Each panel shows an average for a different season. The x-axis is on a log-scale, which means the slope of the curve is related to Kd.

10%) and to a slightly lesser degree by PFT chlorophyll assimilation (~ 6%). The difference in the light field between the two (PFT absorption vs PFT chlorophyll) assimilative runs is caused by: a) the difference in PFT community structure with highly absorbing picophytoplankton being more abundant in PFT absorption assimilation, and low absorbing diatoms being more abundant in PFT chlorophyll assimilation (see Fig.4, Fig.6 and Fig.8); and b) the larger total chlorophyll concentrations in the PFT chlorophyll assimilative run, when compared to the PFT absorption assimilation (Fig.4 and Fig.6). The change in community structure also explains why there is more underwater light in the PFT chlorophyll assimilative run than in the free run (Fig.10), since the smaller phytoplankton size-classes (picophytoplankton and nanophytoplankton) that absorb more photosynthetic energy are more prevalent in the PFT chlorophyll assimilative run than in the free run using the bio-optical module (Fig.4).

It is essential to validate the changes to the underwater light field due to the bio-optical module and assimilation. Since it is hard to get shelf-wide data for PAR, we validated the module skill with the available data for the specific L4 location. Fig.11 compares the PAR attenuation in the water column between the two free runs (pre-existing light scheme vs bio-optical module), the two assimilative runs using the bio-optical module and the L4 observations. For the L4 location we learned (Fig.11) that: 1. as suggested by shelf-wide results the bio-optical module leads to substantially larger light attenuation than the pre-existing light scheme, 2. larger differences in underwater PAR result from the bio-optical module than from the assimilation, 3. the bio-optical module is more skilled to represent the L4 data than the pre-existing light scheme, 4. the module skill in representing PAR attenuation depends largely on the season, with the module doing a much better job in the Spring-Summer period than in Autumn-Winter, 5. overall the light module seems to slightly overestimate PAR attenuation when compared to the observations (especially in the upper 0-30 m), 6. module Kd varies more with depth than the observed Kd. While the results presented in Fig.11 are encouraging and consistent with shelf-wide analysis, it is important to keep in mind the limits of extrapolating general considerations from a location-specific analysis. Furthermore, there were no available L4 data for 2016 phytoplankton, CDOM, POM or sediment, so it is difficult to explain the difference between the PAR field in the module and the observations.

For the shelf-wide model skill we can compare the model and the OC satellite surface visibility (defined as $1/Kd$) for the 490 nm wavelength (the left hand panel of Fig.5). The pre-existing light scheme does not spectrally resolve the underwater light (the visibility in the pre-existing scheme is represented as broadband) and it cannot be directly compared with the 490 nm visibility of the bio-optical module, or the satellite data. However, since broadband light attenuation assumes that the broadband value is sufficiently representative of all the wavelengths from the spectral band, we included the pre-existing scheme into Fig.5 by representing its broadband visibility with a spectrally constant value. It is then shown that for the 490 nm wavelength: a) all simulations tend to underestimate surface visibility and therefore underwater light near the ocean surface, b) the bio-optical module substantially improved the match-ups between free run and satellite visibility at 490 nm, by increasing visibility, c) PFT chlorophyll assimilation using the pre-existing light scheme has little overall capability to improve the near-surface light field, d) both PFT chlorophyll and PFT absorption assimilative runs using the bio-optical module outperform all the other runs in their match-ups with the satellite data. The improved seasonal time series of surface visibility in the two assimilative runs (Fig.5) is related to the improved phytoplankton seasonal dynamics from Fig.6. However the relationship between surface visibility and phytoplankton concentration is not straightforward, as light attenuation includes impact of multiple other constituents (POM, CDOM, sediment).

Additional insight into the model and the satellite surface visibility at 490 nm is provided by two spatial Figures, Fig.12 and Fig.13. Similarly to Fig.5, Fig.12 shows that the pre-existing light module underestimates satellite surface visibility on the NWE Shelf,

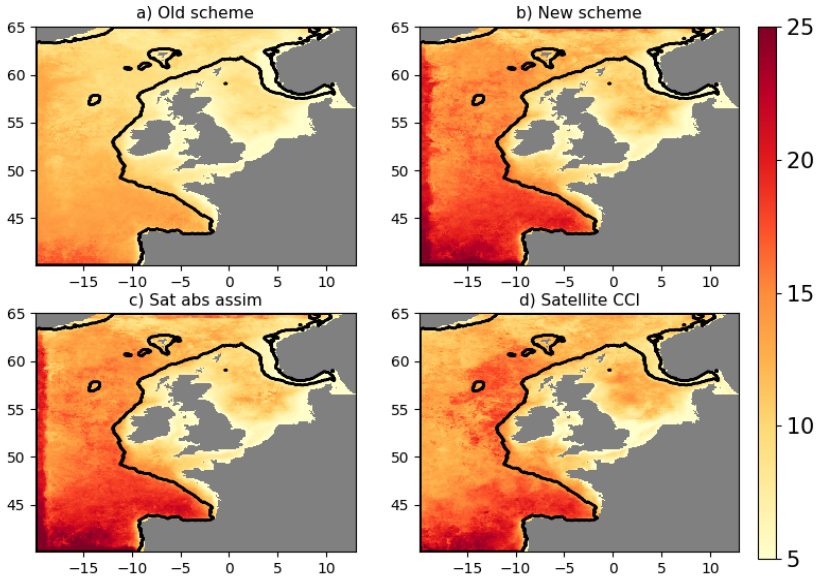


Figure 12. Annual 2016 median distributions of surface visibility (*m*) for the 490 *nm* wavelength, defined as the inverse of Kd at the satellite data locations. The different panels compare a) free run using the pre-existing light scheme ("old scheme"), b) the free run using the bio-optical module ("new scheme"), c) the run assimilating PFT absorption ("sat abs asim") and d) the OC satellite data. The run assimilating the PFT chlorophyll-a into the bio-optical module is not shown, as the distributions are virtually identical to the absorption assimilation (bottom left panel). The full black line marks the boundary of the continental shelf (< 200 m bathymtetry).

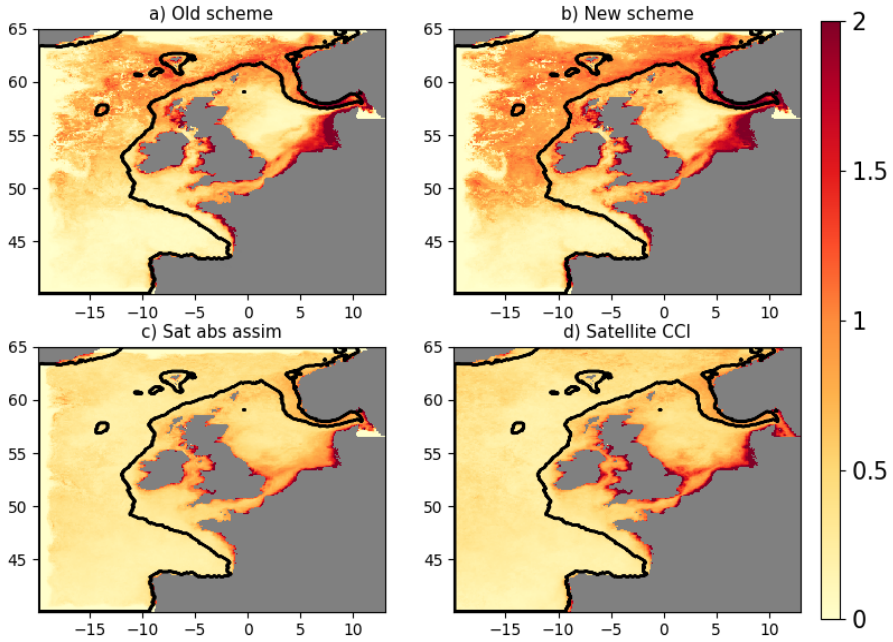


Figure 13. Annual 2016 median distributions of surface total chlorophyll-a (mg/m^3) at the satellite data locations. The different panels compare a) free run using the pre-existing light scheme ("old scheme"), b) the free run using the bio-optical module ("new scheme"), c) the run assimilating PFT absorption ("sat abs asim") and d) the OC satellite data. The run assimilating the PFT chlorophyll-a into the bio-optical module is not shown, as the distributions are virtually identical to the absorption assimilation (bottom left panel).

with substantial improvement carried by the bio-optical module and the two assimilative runs using the bio-optical module. In Fig.13 we show the corresponding total chlorophyll-a surface concentrations, which are (on the NWE Shelf) clearly anti-correlated (Pearson correlation, R=-0.71) with the surface visibility. The relationship between phytoplankton and surface visibility is more visible in the runs using bio-optical module (Fig.12, Fig.13:b-c) as the 490 nm visibility resolved by the bio-optical module is more sensitive to phytoplankton than the broadband visibility of the pre-existing light scheme (Fig.12, Fig.13:a). The spatial analysis from Fig.13 also supports the conclusions derived from the time series presented in Fig.6 and Fig.7: on the NWE Shelf the total chlorophyll-a surface distributions of the assimilative runs match closely with the satellite data.

3.4 General discussion and the future directions

While observations suggest that by introducing the spectrally resolved bio-optical module into ERSEM we improved representation of underwater light field (Fig.5, Fig.11 and Fig.12), it is unclear whether the improved light also improved ERSEM ecosystem dynamics. There is some indication of improvement in model primary productivity (Fig.9), however the overall model skill assessment from Fig.7 shows no significant difference between the performance of the pre-existing light scheme and the bio-optical module. It is particularly disappointing that the spectrally resolved module failed to correct the model phytoplankton seasonal cycle, in particular the rapid and late model Spring bloom (as consistently indicated by both satellite and in situ data, e.g Fig.3). The phytoplankton seasonality is of major importance for ecosystem dynamics as it provides grounds for any higher trophic level processes. Phytoplankon bloom timing and magnitude depend on three key drivers: a) nutrient availability, b) vertical mixing and the c) light availability. More specifically, the onset of Spring bloom is thought to be primarily dependent on the relationship between the depth penetrated by the solar radiation and some effective depth of phytoplankton mixing, which could be the mixed layer depth (as assumed by the critical depth hypothesis of *Sverdrup* [1953]), or some effective depth of turbulent mixing (see the critical turbulence hypothesis, *Huisman et al.* [1999]; *Waniek* [2003]).

Clearly, a model late Spring bloom can be indicative of incorrect model vertical mixing. The sensitivity of primary productivity to the upper ocean mixing scheme is well known (*Oschlies and Garçon* [1999]; *Doney et al.* [2004]) and is often responsible for the deterioration in ecosystem model skill when physical data are assimilated into models (*Berline et al.* [2007]; *Samuelsen et al.* [2009]; *Raghukumar et al.* [2015]). Errors in vertical mixing could potentially delay model Spring blooms, typically when too much mixing reduces phytoplankton concentrations by transporting its biomass to deeper and darker parts in the watercolumn (*Huisman et al.* [1999]; *Taylor and Ferrari* [2011]; *Smyth et al.* [2014]). A promising route to address inconsistencies between the underlying physics (e.g. vertical transport) and biogeochemistry is to use the underwater light field calculated from the bio-optical module to drive temperature in the water column and provide very important feedback from biogeochemistry to physics (*Simonot et al.* [1988]; *Sathyendranath et al.* [1991]; *Edwards et al.* [2004]; *Lengaigne et al.* [2007]; *Zhai et al.* [2011]). Such feedback provides us with a two-way coupled physics-biogeochemical model, which we will introduce in the future within the NEMO-FABM-ERSEM context.

Another possibility for the late model bloom is that ERSEM underestimates the underwater PAR and the environment is too dark for an earlier bloom to kick-in. The Fig.5 indeed suggest that despite of the vanishing module bias in PAR when compared to EUMETSAT POS data, the module overestimates the light attenuation near the water surface. It seems (Fig.12) that the module substantially overestimates light attenuation by some of the represented substances, and this overestimate is large enough to compensate for the POM and sediment backscattering, which was not included into light attenuation within this study. Some further model development might be therefore required in order to include the impact of sediment backscattering, while simultaneously increasing the un-

derwater light field. One important clue might be the mismatch in diatoms and picophytoplankton specific absorption coefficients between the module and the satellite (Fig.8), which points to the large uncertainty in phytoplankton absorption that will need to be addressed in the future. A separate question is the ERSEM PFT community structure that substantially impacts the absorption of light by phytoplankton. The ERSEM PFT community structure is also sensitive to the model parametrization, in particular to the relatively poorly constrained maximum chlorophyll-to-carbon ratio parameters (*Ciavatta et al.* [2014]). However, after all increasing underwater PAR might not have substantial impact on the late phytoplankton bloom, as the pre-existing ERSEM light scheme had 10-20% higher underwater PAR than the module and the phytoplankton seasonal cycle remained broadly unchanged.

The model phytoplankton bloom has also much larger magnitude than what has been observed both in the satellite and in situ data (e.g. Fig.3). This may be explained by the fact that ERSEM substantially overestimates the nitrate concentrations on the NWE Shelf (e.g *Ciavatta et al.* [2018]; *Skákala et al.* [2018], see also Fig.7), providing too much nutrients for the phytoplankton growth. The issue of ERSEM nitrate is unrelated to how the model represents chlorophyll, and improving the ERSEM chlorophyll by assimilation is known to further degrade the nitrate model bias (*Ciavatta et al.* [2018]; *Skákala et al.* [2018]). We anticipate that to improve the ERSEM nitrate concentrations one needs to either focus on the model forcing (river discharge, nitrogen atmospheric deposition), or on some relevant model parameters, such as nitrification rate.

There is also a more general possibility that the current ERSEM version neglects important processes with substantial impact on phytoplankton seasonality. One such process could be dinoflaggelate motility, another process that has been already shown to have positive impact on the simulation of phytoplankton succession is the explicit representation of different xanthophyll photoprotective activities in phytoplankton groups (*Polimene et al.* [2014]). Finally, it is quite possible that to improve ecosystem dynamics under the new module one needs a more complex ERSEM reparametrization. This would not be entirely surprising as the current set of ERSEM parameter values has been chosen to optimize the skill of the model using the pre-existing light scheme and such parametrization might easily become sub-optimal once one introduces large changes to the model. Since model development and parametrization is an arduous task, it is encouraging that it can be partly bypassed by data assimilation, with assimilative runs substantially outperforming free runs in chlorophyll (Fig.6, Fig.7 and Fig.13), underwater light field (Fig.5) and possibly also primary productivity (Fig.9). In fact improvement in our understanding of ecosystem dynamics (e.g carbon cycle, nutrient cycle, trophic export) should be ideally understood as a hand-in-hand effort between the model development and the new types of observational products advancing our assimilation capability. Here the bio-optical module plays an important role in multiple aspects of this process: it improves the model, provides us with a better capacity to assimilate new data into the model and potentially, in the future, it could help to develop, or validate, observational algorithms used to derive biogeochemical fields of interest.

4 Summary

In this work we introduced a novel bio-optical module into an ecosystem model that is used for operational reanalysis and forecasts on the NWE Shelf. The two main advantages of the bio-optical module are that it: a) spectrally resolves the light in the water-column, and b) better accounts for the direction of the light beam. The new module improves the simulation of the underwater light field by providing its spectral decomposition and improving the light attenuation in the water column. The improved representation of the underwater light changes the simulated primary productivity and there is some evidence that the changed primary productivity improves phytoplankton carbon biomass. Much greater improvement in model skill is achieved through assimilating either satel-

lite PFT chlorophyll, or PFT absorption, with both assimilative runs having major positive impact on the model skill to represent chlorophyll seasonal cycle (i.e. the timing and magnitude of Spring bloom), underwater light attenuation and possibly also phytoplankton carbon biomass. The model skill to represent biogeochemical variables is improved dominantly by assimilation, while the model skill to represent underwater light field is improved primarily by the bio-optical module. The importance of the bio-optical module is particularly evident with respect to the currently established assimilation of PFT chlorophyll using the pre-existing ERSEM light scheme, as we have shown that this fails to correct the underwater light field of the free run. We suggest that model simulation of phytoplankton seasonal cycle could be further improved by re-tuning ERSEM parametrization (e.g, addressing the values of phytoplankton specific absorption coefficients), improving nutrient forcing (e.g. river discharge) and improving the underlying physics (e.g vertical mixing). The latter could be potentially addressed by using the bio-optical module to correct temperature profiles in a fully two-way coupled physical-biogeochemical model.

Acknowledgments

This work was carried out as part of the Copernicus Marine Environment Monitoring Service (CMEMS) project OPTIcal data Modelling and Assimilation (OPTIMA) and it supports the CMEMS NOWMAPS project. CMEMS is implemented by MERCATOR OCEAN in the framework of a delegation agreement with the European Union. The work was also supported by the Natural Environment Research Council (NERC) project Combining Autonomous observations and Models for Predicting and Understanding Shelf seas (CAMPUS, contract no NE/R006849/1). The river forcing data were prepared by Sonja van Leeuwen and Helen Powley as part of UK Shelf Seas Biogeochemistry programme (contract no. NE/K001876/1) of the NERC and the Department for Environment Food and Rural Affairs (DEFRA). We thank the European Space Agency Climate Change Initiative "Ocean Colour" (http://www.esa-oceancolour-cci.org) for providing the ocean color data. We acknowledge use of the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, a strategic partnership between the Met Office and the NERC. The ERA-5 atmospheric data used to force the bio-optical module can be freely downloaded from *https://www.ecmwf.int/*, the MODIS data for aerosol optical thickness from *https://modis.gsfc.nasa.gov/data/dataprod*, the NSBC climatological data-set used for model validation can be downladed from *https://icdc.cen.uni-hamburg.de/1/daten/ocean/knsc-hydrographic0/* and the ICES data from *https://www.ices.dk/marine-data/*. The L4 validation data can be obtained from the Western Channel Observatory (*https://www.westernchannelobservatory.org.uk/*). The outputs for the NEMO-FABM-ERSEM simulations are stored on the MONSooN storage facility MASS and can be obtained upon request.

References

Allen, J., and P. Somerfield (2009), A multivariate approach to model skill assessment, *Journal of Marine Systems*, *76*(1-2), 83–94.

Andersson, E. (2003), Modelling the temporal evolution of innovation statistics, *This volume*, pp. 153–164.

Artioli, Y., J. C. Blackford, M. Butenschön, J. T. Holt, S. L. Wakelin, H. Thomas, A. V. Borges, and J. I. Allen (2012), The carbonate system in the north sea: Sensitivity and model validation, *Journal of Marine Systems*, *102*, 1–13.

Baird, M. E., N. Cherukuru, E. Jones, N. Margvelashvili, M. Mongin, K. Oubelkheir, P. J. Ralph, F. Rizwi, B. J. Robson, T. Schroeder, et al. (2016), Remote-sensing reflectance and true colour produced by a coupled hydrodynamic, optical, sediment, biogeochemical model of the great barrier reef, australia: comparison with satellite data, *Environmental modelling & software*, *78*, 79–96.

Baretta, J., W. Ebenhöh, and P. Ruardij (1995), The european regional seas ecosystem model, a complex marine ecosystem model, *Netherlands Journal of Sea Research*, *33*(3-4), 233–246.

Baretta-Bekker, J., J. Baretta, and W. Ebenhöh (1997), Microbial dynamics in the marine ecosystem model ersem ii with decoupled carbon assimilation and nutrient uptake, *Journal of Sea Research*, *38*(3-4), 195–211.

Berline, L., J.-M. Brankart, P. Brasseur, Y. Ourmières, and J. Verron (2007), Improving the physics of a coupled physical–biogeochemical model of the north atlantic through data assimilation: Impact on the ecosystem, *Journal of Marine Systems*, *64*(1-4), 153–172.

Bissett, W., K. Carder, J. Walsh, and D. Dieterle (1999), Carbon cycling in the upper waters of the sargasso sea: Ii. numerical simulation of apparent and inherent optical properties, *Deep Sea Research Part I: Oceanographic Research Papers*, *46*(2), 271–317.

Bissett, W. P., R. Arnone, S. DeBra, D. A. Dieterle, D. Dye, G. J. Kirkpatrick, O. M. Schofield, and G. A. Vargo (2005), Predicting the optical properties of the west florida shelf: resolving the potential impacts of a terrestrial boundary condition on the distribution of colored dissolved and particulate matter, *Marine chemistry*, *95*(3-4), 199–233.

Blackford, J. (1997), An analysis of benthic biological dynamics in a north sea ecosystem model, *Journal of Sea Research*, *38*(3-4), 213–230.

Blackford, J., and F. Gilbert (2007), ph variability and co2 induced acidification in the north sea, *Journal of Marine Systems*, *64*(1-4), 229–241.

Borges, A., L.-S. Schiettecatte, G. Abril, B. Delille, and F. Gazeau (2006), Carbon dioxide in european coastal waters, *Estuarine, Coastal and Shelf Science*, *70*(3), 375–387.

Brewin, R. J., S. Sathyendranath, T. Hirata, S. J. Lavender, R. M. Barciela, and N. J. Hardman-Mountford (2010), A three-component model of phytoplankton size class for the atlantic ocean, *Ecological Modelling*, *221*(11), 1472–1483.

Brewin, R. J., S. Sathyendranath, D. Müller, C. Brockmann, P.-Y. Deschamps, E. Devred, R. Doerffer, N. Fomferra, B. Franz, M. Grant, et al. (2015), The ocean colour climate change initiative: Iii. a round-robin comparison on in-water bio-optical algorithms, *Remote Sensing of Environment*, *162*, 271–294.

Brewin, R. J., S. Ciavatta, S. Sathyendranath, T. Jackson, G. Tilstone, K. Curran, R. L. Airs, D. Cummings, V. Brotas, E. Organelli, et al. (2017), Uncertainty in ocean-color estimates of chlorophyll for phytoplankton groups, *Frontiers in Marine Science*, *4*, 104.

Brewin, R. J., S. Ciavatta, S. Sathyendranath, J. Skákala, J. Bruggeman, D. Ford, and T. Platt (2019), The influence of temperature and community structure on light absorption by phytoplankton in the north atlantic, *Sensors*, *19*(19), 4182.

Bruggeman, J., and K. Bolding (2014), A general framework for aquatic biogeochemical models, *Environmental modelling & software*, *61*, 249–265.

Butenschön, M., J. Clark, J. N. Aldridge, J. I. Allen, Y. Artioli, J. Blackford, J. Bruggeman, P. Cazenave, S. Ciavatta, S. Kay, et al. (2016), Ersem 15.06: a generic model for marine biogeochemistry and the ecosystem dynamics of the lower trophic levels, *Geoscientific Model Development*, *9*(4), 1293–1339.

Campbell, J. W. (1995), The lognormal distribution as a model for bio-optical variability in the sea, *Journal of Geophysical Research: Oceans*, *100*(C7), 13,237–13,254.

Carmillet, V., J.-M. Brankart, P. Brasseur, H. Drange, G. Evensen, and J. Verron (2001), A singular evolutive extended kalman filter to assimilate ocean color data in a coupled physical–biochemical model of the north atlantic ocean, *Ocean Modelling*, *3*(3-4), 167–192.

Ciavatta, S., R. Torres, S. Saux-Picart, and J. I. Allen (2011), Can ocean color assimilation improve biogeochemical hindcasts in shelf seas?, *Journal of Geophysical Research: Oceans*, *116*(C12).

Ciavatta, S., R. Torres, V. Martinez-Vicente, T. Smyth, G. DallâĂŹOlmo, L. Polimene, and J. I. Allen (2014), Assimilation of remotely-sensed optical properties to improve marine biogeochemistry modelling, *Progress in oceanography*, *127*, 74–95.

Ciavatta, S., S. Kay, S. Saux-Picart, M. Butenschön, and J. Allen (2016), Decadal reanalysis of biogeochemical indicators and fluxes in the north west european shelf-sea ecosystem, *Journal of Geophysical Research: Oceans*, *121*(3), 1824–1845.

Ciavatta, S., R. Brewin, J. Skakala, L. Polimene, L. de Mora, Y. Artioli, and J. I. Allen (2018), Assimilation of ocean-color plankton functional types to improve marine ecosystem simulations, *Journal of Geophysical Research: Oceans*, *123*(2), 834–854.

Ciavatta, S., S. Kay, R. Brewin, R. Cox, A. Di Cicco, F. Nencioli, L. Polimene, M. Sammartino, R. Santoleri, J. Skákala, et al. (2019), Ecoregions in the mediterranean sea through the reanalysis of phytoplankton functional types and carbon fluxes, *Journal of Geophysical Research: Oceans*.

Cummings, J., L. Bertino, P. Brasseur, I. Fukumori, M. Kamachi, M. J. Martin, K. Mogensen, P. Oke, C. E. Testut, J. Verron, et al. (2009), Ocean data assimilation systems for godae, *Oceanography*, *22*(3), 96–109.

de Haas, H., T. C. van Weering, and H. de Stigter (2002), Organic carbon in shelf seas: sinks or sources, processes and products, *Continental Shelf Research*, *22*(5), 691–717.

De Mora, L., M. Butenschon, and J. Allen (2013), How should sparse marine in situ measurements be compared to a continuous model: an example, *Geoscientific Model Development*, *6*(2), 533–548.

De Mora, L., M. Butenschon, and J. Allen (2016), The assessment of a global marine ecosystem model on the basis of emergent properties and ecosystem function: a case study with ersem, *Geosci. Model Dev.*, *9*, 59–76.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(613), 3385–3396.

Dickey, T. D., G. W. Kattawar, and K. J. Voss (2011), Shedding new light on light in the ocean, *Phys. Today*, *64*(4), 44–49.

Doney, S. C., K. Lindsay, K. Caldeira, J.-M. Campin, H. Drange, J.-C. Dutay, M. Follows, Y. Gao, A. Gnanadesikan, N. Gruber, et al. (2004), Evaluating global ocean carbon models: The importance of realistic physics, *Global Biogeochemical Cycles*, *18*(3).

Doney, S. C., K. Lindsay, I. Fung, and J. John (2006), Natural variability in a stable, 1000-yr global coupled climate–carbon cycle simulation, *Journal of climate*, *19*(13), 3033–3054.

Dutkiewicz, S., A. Hickman, O. Jahn, W. Gregg, C. Mouw, and M. Follows (2015), Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model, *Biogeosciences*, *12*(14), 4447–4481.

Edwards, A. M., D. G. Wright, and T. Platt (2004), Biological heating effect of a band of phytoplankton, *Journal of Marine Systems*, *49*(1-4), 89–103.

Edwards, C. A., A. M. Moore, I. Hoteit, and B. D. Cornuelle (2015), Regional ocean data assimilation, *Annual review of marine science*, *7*, 21–42.

Edwards, K., R. Barciela, and M. Butenschön (2012), Validation of the nemo-ersem operational ecosystem model for the north west european continental shelf, *Ocean Science*, *8*(6), 983–1000.

Finenko, Z., N. Hoepffner, R. Williams, and S. Piontkovski (2003), Phytoplankton carbon to chlorophyll a ratio: response to light, temperature and nutrient limitation.

Fontana, C., C. Grenz, and C. Pinazo (2010), Sequential assimilation of a year-long time-series of seawifs chlorophyll data into a 3d biogeochemical model on the french mediterranean coast, *Continental Shelf Research*, *30*(16), 1761–1771.

Ford, D., and R. Barciela (2017), Global marine biogeochemical reanalyses assimilating two different sets of merged ocean colour products, *Remote Sensing of Environment*, *203*, 40–54.

Ford, D., K. Edwards, D. Lea, R. Barciela, M. Martin, and J. Demaria (2012), Assimilating globcolour ocean colour data into a pre-operational physical-biogeochemical model, *Ocean Science*, *8*(5), 751–771.

Ford, D. A., J. van der Molen, K. Hyder, J. Bacon, R. Barciela, V. Creach, R. McEwan, P. Ruardij, and R. Forster (2017), Observing and modelling phytoplankton community structure in the north sea, *Biogeosciences*, *14*(6), 1419–1444.

Garcia, H. E., R. A. Locarnini, T. P. Boyer, J. I. Antonov, O. K. Baranova, M. M. Zweng, J. R. Reagan, D. R. Johnson, A. V. Mishonov, and S. Levitus (2013), World ocean atlas 2013. volume 4, dissolved inorganic nutrients (phosphate, nitrate, silicate).

Gehlen, M., R. Barciela, L. Bertino, P. Brasseur, M. Butenschön, F. Chai, A. Crise, Y. Drillet, D. Ford, D. Lavoie, et al. (2015), Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, *Journal of Operational Oceanography*, *8*(sup1), s168–s187.

Geider, R., H. MacIntyre, and T. Kana (1997), Dynamic model of phytoplankton growth and acclimation: responses of the balanced growth rate and the chlorophyll a: carbon ratio to light, nutrient-limitation and temperature, *Marine Ecology Progress Series*, *148*, 187–200.

Gregg, W. W. (2002), *A coupled ocean-atmosphere radiative model for global ocean biogeochemical models*, National Aeronautics and Space Administration, Goddard Space Flight Center.

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.

Gregg, W. W., and N. W. Casey (2007), Modeling coccolithophores in the global oceans, *Deep Sea Research Part II: Topical Studies in Oceanography*, *54*(5-7), 447–477.

Gregg, W. W., and N. W. Casey (2009), Skill assessment of a spectral ocean–atmosphere radiative model, *Journal of Marine Systems*, *76*(1-2), 49–63.

Gregg, W. W., and C. S. Rousseaux (2016), Directional and spectral irradiance in ocean models: effects on simulated global phytoplankton, nutrients, and primary production, *Frontiers in Marine Science*, *3*, 240.

Gregg, W. W., and C. S. Rousseaux (2017), Simulating pace global ocean radiances, *Frontiers in Marine Science*, *4*, 60.

Gregg, W. W., P. Ginoux, P. S. Schopf, and N. W. Casey (2003), Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model, *Deep Sea Research Part II: Topical Studies in Oceanography*, *50*(22-26), 3143–3169.

Groom, S. B., S. Sathyendranath, Y. Ban, S. Bernard, B. Brewin, V. Brotas, C. Brockmann, P. Chauhan, J.-k. Choi, A. Chuprin, et al. (2019), Satellite ocean colour: current status and future perspective, *Frontiers in Marine Science*, *6*, 485.

Henson, S. A., J. L. Sarmiento, J. P. Dunne, L. Bopp, I. D. Lima, S. C. Doney, J. G. John, and C. Beaulieu (2010), Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity.

Hinrichs, I., V. Gouretski, J. Pätsch, K.-C. Emeis, and D. Stammer (2017), North sea biogeochemical climatology (version 1.1).

Holt, J., M. Butenschon, S. Wakelin, Y. Artioli, and J. Allen (2012), Oceanic controls on the primary production of the northwest european continental shelf: model experiments under recent past conditions and a potential future scenario, *Biogeosciences*, *9*, 97–117.

Hoteit, I., G. Triantafyllou, and G. Petihakis (2005), Efficient data assimilation into a complex, 3-d physical-biogeochemical model using partially-local kalman filters.

Huisman, J., P. van Oostveen, and F. J. Weissing (1999), Critical depth and critical turbulence: two different mechanisms for the development of phytoplankton blooms, *Limnology and oceanography*, *44*(7), 1781–1787.

Ishizaka, J. (1990), Coupling of coastal zone color scanner data to a physical-biological model of the southeastern us continental shelf ecosystem: 2. an eulerian model, *Journal of Geophysical Research: Oceans*, *95*(C11), 20,183–20,199.

Jackson, T., S. Sathyendranath, and F. Mélin (2017), An improved optical classification scheme for the ocean colour essential climate variable and its applications, *Remote Sensing of Environment*, *203*, 152–161.

Jahnke, R. A. (2010), Global synthesis, in *Carbon and nutrient fluxes in continental margins*, pp. 597–615, Springer.

Jiang, M.-S., F. Chai, R. Dugdale, F. Wilkerson, T.-H. Peng, and R. Barber (2003), A nitrate and silicate budget in the equatorial pacific ocean: a coupled physical–biological model study, *Deep Sea Research Part II: Topical Studies in Oceanography*, *50*(22-26), 2971–2996.

Jones, E. M., M. E. Baird, M. Mongin, J. Parslow, J. Skerratt, J. Lovell, N. Margvelashvili, R. J. Matear, K. Wild-Allen, B. Robson, 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.

Kalaroni, S., K. Tsiaras, G. Petihakis, I. Hoteit, A. Economou-Amilli, and G. Triantafyllou (2016), Data assimilation of depth-distributed satellite chlorophyll-α in two mediterranean contrasting sites, *Journal of Marine Systems*, *160*, 40–53.

Kalnay, E. (2003), *Atmospheric modeling, data assimilation and predictability*, Cambridge university press.

Key, R. M., A. Olsen, S. van Heuven, S. K. Lauvset, A. Velo, X. Lin, C. Schirnick, A. Kozyr, T. Tanhua, M. Hoppema, et al. (2015), Global ocean data analysis project, version 2 (glodapv2).

King, R. R., J. While, M. J. Martin, D. J. Lea, B. Lemieux-Dudon, J. Waters, and E. OâĂŹDea (2018), Improving the initialisation of the met office operational shelf-seas model, *Ocean Modelling*, *130*, 1–14.

Laufkötter, C., M. Vogt, and N. Gruber (2013), Long-term trends in ocean plankton production and particle export between 1960-2006, *Biogeosciences*, *10*(11), 7373–7393.

Lauvset, S. K., R. M. Key, A. Olsen, S. van Heuven, A. Velo, X. Lin, C. Schirnick, A. Kozyr, T. Tanhua, M. Hoppema, et al. (2016), A new global interior ocean mapped climatology: The 1× 1 glodap version 2, *Earth System Science Data*, *8*, 325–340.

Lee, Z.-P., K.-P. Du, and R. Arnone (2005), A model for the diffuse attenuation coefficient of downwelling irradiance, *Journal of Geophysical Research: Oceans*, *110*(C2).

Lengaigne, M., C. Menkes, O. Aumont, T. Gorgues, L. Bopp, J.-M. André, and G. Madec (2007), Influence of the oceanic biology on the tropical pacific climate in a coupled general circulation model, *Climate Dynamics*, *28*(5), 503–516.

Lenhart, H.-J., D. K. Mills, H. Baretta-Bekker, S. M. Van Leeuwen, J. Van Der Molen, J. W. Baretta, M. Blaas, X. Desmit, W. Kühn, G. Lacroix, et al. (2010), Predicting the consequences of nutrient reduction on the eutrophication status of the north sea, *Journal of Marine Systems*, *81*(1-2), 148–170.

MacLachlan, C., A. Arribas, K. Peterson, A. Maidens, D. Fereday, A. Scaife, M. Gordon, M. Vellinga, A. Williams, R. Comer, et al. (2015), Global seasonal forecast system version 5 (glosea5): a high-resolution seasonal forecast system, *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1072–1084.

Madec, G., et al. (2015), Nemo ocean engine.

Maier-Reimer, E., I. Kriest, J. Segschneider, and P. Wetzel (2005), The hamburg ocean carbon cycle model hamocc5. 1-technical description release 1.1.

Manizza, M., C. Le Quéré, A. J. Watson, and E. T. Buitenhuis (2005), Bio-optical feedbacks among phytoplankton, upper ocean physics and sea-ice in a global model, *Geophysical Research Letters*, *32*(5).

Marinov, I., S. C. Doney, and I. D. Lima (2010), Response of ocean phytoplankton community structure to climate change over the 21st century: partitioning the effects of nutrients, temperature and light.

Mobley, C., L. Sundman, W. Bissett, and B. Cahill (2009), Fast and accurate irradiance calculations for ecosystem models, *Biogeosciences Discussions*, *6*(6), 10,625–10,662.

Mogensen, K., M. Balmaseda, A. Weaver, M. Martin, and A. Vidard (2009), Nemovar: A variational data assimilation system for the nemo ocean model, *ECMWF newsletter*, *120*, 17–22.

Mogensen, K., M. A. Balmaseda, A. Weaver, et al. (2012), The nemovar ocean data assimilation system as implemented in the ecmwf ocean analysis for system 4.

Moore, T. S., J. W. Campbell, and M. D. Dowell (2009), A class-based approach to characterizing and mapping the uncertainty of the modis ocean chlorophyll product, *Remote Sensing of Environment*, *113*(11), 2424–2430.

Natvik, L.-J., and G. Evensen (2003), Assimilation of ocean colour data into a biochemical model of the north atlantic: Part 1. data assimilation experiments, *Journal of Marine Systems*, *40*, 127–153.

Nerger, L., and W. W. Gregg (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.

Nerger, L., and W. W. Gregg (2008), Improving assimilation of seawifs data by the application of bias correction with a local seik filter, *Journal of marine systems*, *73*(1-2), 87–102.

O'Dea, E., R. Furner, S. Wakelin, J. Siddorn, J. While, P. Sykes, R. King, J. Holt, and H. Hewitt (2017), The co5 configuration of the 7 km atlantic margin model: large-scale biases and sensitivity to forcing, physics options and vertical resolution, *Geoscientific Model Development*, *10*(8), 2947.

Oschlies, A., and V. Garçon (1999), An eddy-permitting coupled physical-biological model of the north atlantic: 1. sensitivity to advection numerics and mixed layer physics, *Global Biogeochemical Cycles*, *13*(1), 135–160.

Palmer, J., and I. Totterdell (2001), Production and export in a global ocean ecosystem model, *Deep Sea Research Part I: Oceanographic Research Papers*, *48*(5), 1169–1198.

Pauly, D., V. Christensen, S. Guénette, T. J. Pitcher, U. R. Sumaila, C. J. Walters, R. Watson, and D. Zeller (2002), Towards sustainability in world fisheries, *Nature*, *418*(6898), 689.

Polimene, L., S. D. Archer, M. Butenschön, and J. I. Allen (2012), A mechanistic explanation of the sargasso sea dms âĂIJsummer paradoxâĂİ, *Biogeochemistry*, *110*(1-3), 243–255.

Polimene, L., C. Brunet, M. Butenschön, V. Martinez-Vicente, C. Widdicombe, R. Torres, and J. Allen (2014), Modelling a light-driven phytoplankton succession, *Journal of plankton research*, *36*(1), 214–229.

Pradhan, H. K., C. Völker, S. Losa, A. Bracher, and L. Nerger (2019), Assimilation of global total chlorophyll oc-cci data and its impact on individual phytoplankton fields, *Journal of Geophysical Research: Oceans*, *124*(1), 470–490.

Raghukumar, K., C. A. Edwards, N. L. Goebel, G. Broquet, M. Veneziani, A. M. Moore, and J. P. Zehr (2015), Impact of assimilating physical oceanographic data on modeled ecosystem dynamics in the california current system, *Progress in Oceanography*, *138*, 546–558.

Reynolds, R. W., T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax (2007), Daily high-resolution-blended analyses for sea surface temperature, *Journal of Climate*, *20*(22), 5473–5496.

Samuelsen, A., L. Bertino, and C. Hansen (2009), Impact of data assimilation of physical variables on the spring bloom from topaz operational runs in the north atlantic, *Ocean Science*, *5*(4), 635–647.

Sathyendranath, S., A. D. Gouveia, S. R. Shetye, P. Ravindran, and T. Platt (1991), Biological control of surface temperature in the arabian sea, *Nature*, *349*(6304), 54.

Sathyendranath, S., R. Brewin, C. Brockmann, V. Brotas, S. Ciavatta, A. Chuprin, A. Couto, R. Doerffer, M. Dowell, M. Grant, et al. (2016), Creating an ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change initiative, *Remote Sens. Environ.*

Sathyendranath, S., R. J. Brewin, C. Brockmann, V. Brotas, B. Calton, A. Chuprin, P. Cipollini, A. B. Couto, J. Dingle, R. Doerffer, et al. (2019), An ocean-colour time series for use in climate studies: The experience of the ocean-colour climate change

initiative (oc-cci), *Sensors*, *19*(19), 4285.

Schutgens, N., S. Tsyro, E. Gryspeerdt, D. Goto, N. Weigum, M. Schulz, and P. Stier (2017), On the spatio-temporal representativeness of observations, *Atmospheric Chemistry and Physics Discussions*.

Shulman, I., S. Frolov, S. Anderson, B. Penta, R. Gould, P. Sakalaukus, and S. Ladner (2013), Impact of bio-optical data assimilation on short-term coupled physical, bio-optical model predictions, *Journal of Geophysical Research: Oceans*, *118*(4), 2215–2230.

Simonot, J.-y., E. Dollinger, and H. Le Treut (1988), Thermodynamic-biological-optical coupling in the oceanic mixed layer, *Journal of Geophysical Research: Oceans*, *93*(C7), 8193–8202.

Skákala, J., D. Ford, R. J. Brewin, R. McEwan, S. Kay, B. Taylor, L. de Mora, and S. Ciavatta (2018), The assimilation of phytoplankton functional types for operational forecasting in the northwest european shelf, *Journal of Geophysical Research: Oceans*, *123*(8), 5230–5247.

Smyth, T. J., and Y. Artioli (2010), Global inherent optical properties from SeaWiFS data, doi:10.1594/PANGAEA.741913.

Smyth, T. J., G. F. Moore, T. Hirata, and J. Aiken (2006), Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment, *Applied Optics*, *45*(31), 8116–8131.

Smyth, T. J., I. Allen, A. Atkinson, J. T. Bruun, R. A. Harmer, R. D. Pingree, C. E. Widdicombe, and P. J. Somerfield (2014), Ocean net heat flux influences seasonal to interannual patterns of plankton abundance, *PloS one*, *9*(6).

Sverdrup, H. (1953), On conditions for the vernal blooming of phytoplankton, *J. Cons. Int. Explor. Mer*, *18*(3), 287–295.

Taylor, J. R., and R. Ferrari (2011), Shutdown of turbulent convection as a new criterion for the onset of spring phytoplankton blooms, *Limnology and Oceanography*, *56*(6), 2293–2307.

Torres, R., J. Allen, and F. Figueiras (2006), Sequential data assimilation in an upwelling influenced estuary, *Journal of Marine Systems*, *60*(3-4), 317–329.

Wakelin, S., J. Holt, J. Blackford, J. Allen, M. Butenschön, and Y. Artioli (2012), Modeling the carbon fluxes of the northwest european continental shelf: Validation and budgets, *Journal of Geophysical Research: Oceans*, *117*(C5).

Waniek, J. J. (2003), The role of physical forcing in initiation of spring blooms in the northeast atlantic, *Journal of Marine Systems*, *39*(1-2), 57–82.

Waters, J., D. J. Lea, M. J. Martin, I. Mirouze, A. Weaver, and J. While (2015), Implementing a variational data assimilation system in an operational 1/4 degree global ocean model, *Quarterly Journal of the Royal Meteorological Society*, *141*(687), 333–349.

Widdicombe, C., D. Eloire, D. Harbour, R. Harris, and P. Somerfield (2010), Long-term phytoplankton community dynamics in the western english channel, *Journal of Plankton Research*, *32*(5), 643–655.

Woodham, R. (2011), Defence applications of operational oceanography, in *Operational oceanography in the 21st century*, pp. 659–679, Springer.

Xiu, P., and F. Chai (2014), Connections between physical, optical and biogeochemical processes in the pacific ocean, *Progress in Oceanography*, *122*, 30–53.

Zhai, L., C. Tang, T. Platt, and S. Sathyendranath (2011), Ocean response to attenuation of visible light by phytoplankton in the gulf of st. lawrence, *Journal of Marine Systems*, *88*(2), 285–297.

Zhao, J., B. Barnes, N. Melo, D. English, B. Lapointe, F. Muller-Karger, B. Schaeffer, and C. Hu (2013), Assessment of satellite-derived diffuse attenuation coefficients and euphotic depths in south florida coastal waters, *Remote Sensing of Environment*, *131*, 38–50.

Zielinski, O., O. Llinás, A. Oschlies, and R. Reuter (2002), Underwater light field and its effect on a one-dimensional ecosystem model at station estoc, north of the canary is-

lands, *Deep Sea Research Part II: Topical Studies in Oceanography*, *49*(17), 3529–3542.