PHYTOplankton biomass and diversity Climate Change Initiative (PHYTO-CCI)

ESA 4000147645/25/I-LR
D2.2 Algorithm Theoretical Baseline Document (ATBD)

Project led by

PML | Plymouth Marine Laboratory

In cooperation with

Document properties

Project:	PHYTO-CCI
Document title:	Algorithm Theoretical Baseline Document (ATBD)
Reference:	D2.2
Date:	17/06/2026
Version:	1.0

Science leads: Gemma Kulk & Victor Martinez-Vicente – PML

Reviewed: Lianne Jackson – PML

Sarah Connors – ESA

Digitally signed by Sarah Connors
Date: 2026.07.09 11:25:23 +01'00'

Address: Plymouth Marine Laboratory (PML)
Prospect Place
The Hoe
Plymouth
PL1 2DH
United Kingdom
Tel: +44 (0)1752 633100

Document version history

Version	Date	Author(s) / Contributor(s)	Comment
0.0	17/12/2025	Gemma Kulk	Template
1.0 (draft)	17/06/2026	Gemma Kulk, Marco Bellacicco, Astrid Bracher, Hongyan Xi, Victor Martinez-Vicente	First draft
1.0	17/06/2026	Gemma Kulk, Marco Bellacicco, Astrid Bracher, Hongyan Xi, Victor Martinez-Vicente	First release

Index

1 Introduction

1.1 Purpose of document

This Algorithm Theoretical Baseline Document (ATBD) describes the satellite-retrieval algorithms of phytoplankton carbon biomass and pigment diversity that are used by the PHYTOplankton biomass and diversity Climate Change Initiative (PHYTO-CCI). The satellite retrieval algorithms described in this document are based on previous developments by the scientific community, and references are used throughout the document to highlight these earlier results. While different approaches were followed for each algorithm (i.e., (semi-)analytical, empirical and machine learning models), care has been taken that the input data to the different algorithms is consistent and includes climate quality datasets from the ESA Climate Change Initiative.

1.2 Structure

The algorithms for phytoplankton carbon are described in Section 2 and those for phytoplankton pigments in Section 3. For each algorithm, a brief introduction is provided, and the algorithm itself is described, detailing the input and output data, the algorithm and the quality of the data products. In Section 4, the approach to algorithm blending is described. The references are provided in Section 5.

1.3 Related documents

Documents related to the PHYTO-CCI ATBD are the PHYTO-CCI Product Validation and Algorithm Selection Report (PVASR), End-to-End ECV Uncertainty Budget (E3UB) and Algorithm Development Plan (ADP).

2 Phytoplankton carbon biomass

2.1 Algorithm A

2.1.1 Description

The first satellite-based algorithm for phytoplankton carbon is a particle backscattering-based empirical algorithm based on the original approach of Behrenfeld *et al.* (2005) with refinements developed by Bellacicco *et al.* (2016, 2018, 2019, 2020) including recent updates from Yang *et al.* (2024). The algorithm of Behrenfeld *et al.* (2005) was originally developed to provide an alternative proxy of phytoplankton biomass (i.e., carbon instead of chlorophyll-a) and is based on the light backscattering coefficient in seawater due to particles ($b_{bp}(\lambda)$) and the contribution of non-algal particles (NAP) to b_{bp} at 443 nm, assuming a constant value of 0.00035 m^{-1}. Bellacicco *et al.* (2016, 2018, 2019, 2020) demonstrated that this coefficient (i.e., $b_{bNAP}(443)$) varies in space and time, and with depth, highlighting the importance in considering its variability for a more reliable estimate of phytoplankton carbon and description of the marine ecosystem. Yet, Bellacicco *et al.* (2020) also reported the intrinsic caveats of the method, highlighting the reduced reliability of retrievals in the ocean subtropical gyres where the assumption that chlorophyll-a and particle backscattering covary does not hold and photoacclimation drives chlorophyll-a dynamics instead of biomass. Yang *et al.* (2024) therefore refined the method of Bellacicco *et al.* (2020), correcting the physiological effects on the chlorophyll-a and particle backscattering relationship that is at the base of the phytoplankton carbon computation, resulting in a substantial improvement of phytoplankton carbon retrieval in the subtropical ocean gyres.

2.1.2 Algorithm definition

Input data

The following input data are required for the satellite-based phytoplankton carbon algorithm which is based on recent updates as reported in Yang *et al.* (2024) and Volta *et al.* (2025):

- Chlorophyll-a concentration from OC-CCI version 6.0
- Particle backscattering coefficient at 443 nm from OC-CCI version 6.0
- Annual climatology of backscattering of non-algal particles developed by Volta *et al.* (2025)

It is noted that the particle backscattering of OC-CCI is derived using the Quasi-Analytical Algorithm of Lee *et al.* (2002). The climatology of backscattering of non-algal particles is derived from chlorophyll-a, the backscattering coefficient at 443 nm and the attenuation coefficient at 490 nm from OC-CCI version 6, sea surface temperature from SST-CCI version 2.1, photosynthetically active radiation from GlobColour and mixed layer depth from the Global ocean Reanalysis Ensemble Product available from the Copernicus Marine Service, with full details described in Volta *et al.* (2025).

All data are required at 4 km spatial and monthly temporal resolution. The annual climatology of backscattering of non-algal particles is available at 25 km spatial resolution and was regridded to 4 km resolution to match the grid of the other input data.

Retrieval algorithm

The refined phytoplankton carbon algorithm of Yang *et al.* (2024), recently applied in Volta *et al.* (2005), estimates the background non-algal particle backscattering (b_{bNAP}) by embedding the Behrenfeld *et al.* (2005) photoacclimation framework into a constrained non-linear inversion. The algorithm starts from the photoacclimation model linking chlorophyll-a (B) to particle backscattering (b_{bp}):

$$Chl = C \times \theta \text{ , where} \quad (2.1.1)$$

$$C = 13{,}000 \times \left(b_{bp} - b_{bNAP}\right) \text{ , and} \quad (2.1.2)$$

$$\theta = \left(\theta_{min} + (\theta_{max} - \theta_{min}) \times \mathrm{e}^{-3E_g}\right) . \quad (2.1.3)$$

Equation 2.1.2 describes phytoplankton carbon biomass (C) as a function of particle backscattering (b_{bp}) and background backscattering from non-algal particles (b_{bNAP}), where 13,000 represents the conversion factor between b_{bp} and C (Behrenfeld *et al.*, 2005), while Equation 2.1.3 describes the photoacclimation component of the model with the chlorophyll-to-carbon ratio (θ) as a function of the minimum (θ_{min}) and maximum ratio (θ_{max}) and the median light level in the mixed layer (E_g). Here, the θ_{min} accounts for the nutrient limitation on θ under saturated light conditions, decreasing with increasing nutrient stress (Geider *et al.*, 1998; Laws and Bannister, 1980), and θ_{max} accounts for the temperature effect on θ, increasing with temperature (Geider, 1987). In Equation 2.1.1-3, the b_{bNAP}, θ_{min} and θ_{max} are unknown and Yang *et al.* (2024) retrieved these unknowns simultaneously using non-linear least-squares regression. The inversion is stabilised by prescribing θ_{max} through the empirical temperature-dependent formulation of Geider (1987):

$$\theta_{max} = \frac{1}{43.4 - 1.14 \times SST} , \quad (2.1.4)$$

where SST is the sea surface temperature. This reduces the number of free parameters and avoids overfitting of the algorithm in oligotrophic regions. With θ_{max} fixed, the model retrieves b_{bNAP} and θ_{min} simultaneously using non-linear least-squares regression applied to Chl, b_{bp}, SST and E_g fields.

The regression isolates the fraction of b_{bp} that does not covary with phytoplankton carbon biomass, i.e., b_{bNAP} by fitting the observed $Chl - b_{bp}$ relationship while explicitly accounting for photoacclimation-driven variability in θ. This approach overcomes the failure of linear methods in subtropical gyres, where photoacclimation dominates chlorophyll-a variability, and yields spatially coherent and physically consistent b_{bNAP} estimates suitable for retrieval of phytoplankton carbon at the global scale. For more details see Behrenfeld *et al.* (2005), Bellacicco *et al.* (2016, 2018, 2019, 2020), Yang *et al.* (2024) and Volta *et al.* (2025).

Output data

The satellite-derived phytoplankton carbon concentration (in mg C m^{-3}) is available at 4 km spatial resolution for each month between 1997 and 2025. The algorithm allows to detect the total phytoplankton carbon biomass, with no specific distinction between different phytoplankton size classes. The algorithm of Yang *et al.* (2024) showed better global estimation of phytoplankton carbon compared with the original method of Behrenfeld *et al.* (2005), which substantially overestimates the phytoplankton carbon biomass estimated from satellite observations, especially in the subtropical gyres where chlorophyll-a does not covary with particle backscattering (Figure 1).

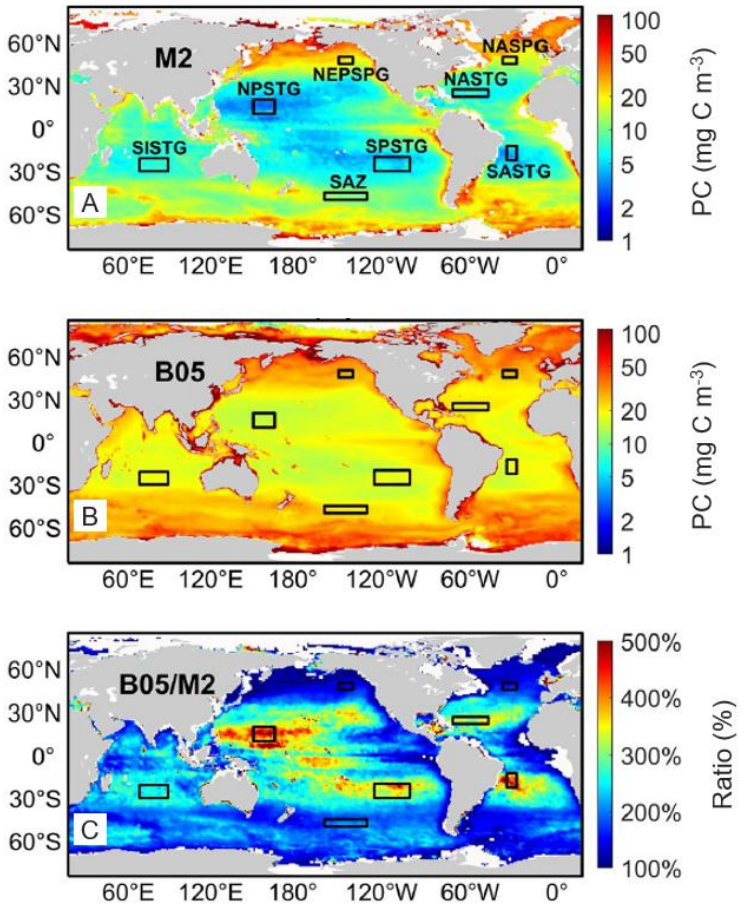

Figure 1. Annual climatology of phytoplankton carbon (PC) derived by A) Yang et al. *(2024), B) Behrenfeld* et al. *(2005) and C) their ratio (B05/M2). The squares respectively represent five subtropical gyres (North Atlantic subtropical gyre, NASTG; North Pacific subtropical gyre, NPSTG; South Atlantic subtropical gyre, SASTG; South Pacific subtropical gyre, SPSTG; South Indian subtropical gyre, SISTG) defined by Morel* et al. *(2010) as well as North Atlantic sub-polar gyre (NASPG), Northeast Pacific sub-polar gyre (NEPSPG) and Southern Ocean Sub-Antarctic Zone (SAZ).*

2.1.3 Quality assessment

Yang *et al.* (2024) validated their phytoplankton carbon algorithm using the *in situ* database collated by Martinez-Vicente *et al.* (2017), which contains observations of picophytoplankton carbon from flow cytometry measurements in primarily the Atlantic Ocean. They showed that the statistical performance of their algorithm exceeds that of previously described particle backscattering-based empirical algorithms (Behrenfeld *et al.*, 2005; Bellacicco *et al.*, 2018), with a lower total error and a higher correlation coefficient (Figure 2). The decomposition of the total error showed that the error mainly comes from the variance between the two datasets used for the validation (i.e., *in situ* and satellite) rather than from systematic biases. However, the relative system error indicates that the algorithm of Yang *et al.* (2024) provides a relatively weak overestimation of phytoplankton carbon (7%). It is also noted that the performance of the algorithm in estimating phytoplankton carbon in the subtropical gyres is satisfying, with a total error of 6.4 mg m^{-3}, a system relative error of 20% and an overall relative error of 83%. The limitations and caveats of particle backscattering-based empirical algorithms of phytoplankton carbon are documented in more detail in Yang *et al.* (2024), Volta *et al.* (2025) and previous publications by Bellacicco *et al.* (2016, 2018, 2020).

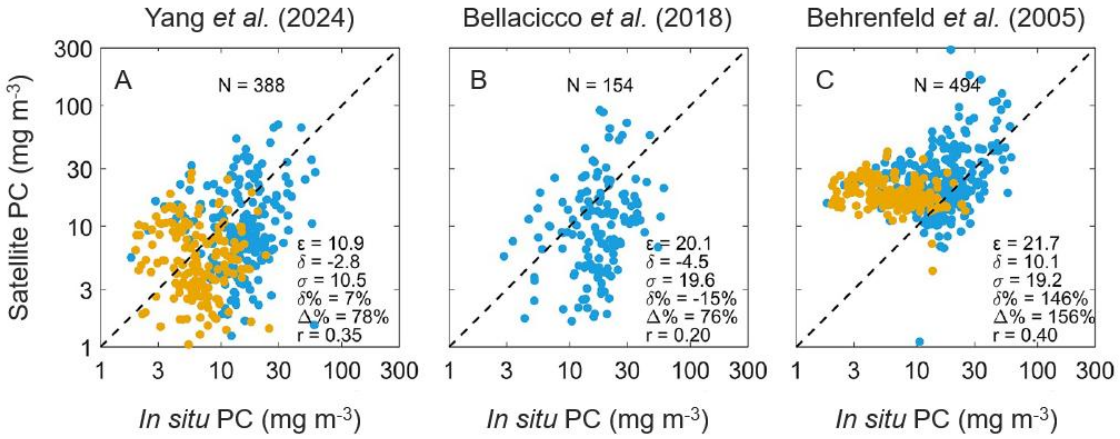

Figure 2. Validation of satellite-based phytoplankton carbon (PC) with in situ *observations with A) the algorithm of Yang* et al. (2024), i.e., algorithm A in PHYTO-CCI, B) the algorithm of Bellacicco et al. *(2018) which assumes no covariation between chlorophyll-a and backscatter and C) the original algorithm of Behrenfeld* et al. *(2005). The blue and orange dots represent match-up data points outside and inside gyres, respectively. Figure adjusted from Yang* et al. *(2024).*

2.2 Algorithm B

2.2.1 Description

The second algorithm is a particle backscattering-based allometric semi-analytical model to derive the slope and intercept of the phytoplankton particle size distribution (PSD) from which concentrations of chlorophyll-a, phytoplankton carbon and particulate organic carbon can be derived (Kostadinov *et al.*, 2009, 2016, 2023). It is based around the idea that the PSD of phytoplankton is a key property affecting both optical properties and cellular physiological and biogeochemical properties, i.e., linking ocean colour remote sensing and ecosystem/biogeochemical characteristics (Kostadinov *et al.*, 2023). Kostadinov *et al.* (2009) first developed an algorithm to derive phytoplankton PSDs using the spectral shape of remote sensing reflectances, assuming a single population of particles that represented phytoplankton cells as well as non-algal particles. Subsequently, the retrieved slope and intercept of the PSD allowed for the quantification of absolute and fractional phytoplankton size classes (i.e., pico-, nano- and microphytoplankton) based on bio-volume (Kostadinov *et al.*, 2010) or phytoplankton carbon (Kostadinov *et al.*, 2016) via allometric relationships (Menden-Deuer and Lessard, 2000). The latest development of the PSD algorithm saw

a major improvement in the representation of particles with two separate particle populations being modelled, i.e., living phytoplankton cells and non-algal particles (Kostadinov *et al.*, 2023). Here we describe this latest version of the PSD algorithm by Kostadinov *et al.* (2023), from which phytoplankton carbon can be estimated for the total phytoplankton community and three specific size classes, namely pico-, nano- and microphytoplankton.

2.2.2 Algorithm definition

Input data

The following input data are required for the satellite-based phytoplankton carbon algorithm:

- Remote sensing reflectances from OC-CCI version 6.0

All data are required at 4 km spatial and monthly temporal resolution.

Retrieval algorithm

The phytoplankton carbon algorithm of Kostadinov *et al.* (2023) is based on the retrieval of the Particle Size Distribution (PSD) of suspended particles in near-surface waters. The forward optical model considers two distinct particle populations: phytoplankton and non-algal particles (NAPs). Phytoplankton are modelled as coasted spheres following the Equivalent Algal Populations (EAP) framework and NAPs are modelled as homogenous spheres. The forward model uses Aden-Kerker and Mie scattering computations, for coated and homogenous spheres, respectively, to model the total particulate spectral backscattering coefficient as the sum of phytoplankton and NAP backscattering. The PSD retrieval is achieved via spectral angle mapping (SAM), which uses backscattering end-members created by the forward model. The PSD is then used to retrieve the size-partitioned absolute and fractional phytoplankton carbon concentrations using allometric coefficients. The algorithm is described in detail by Kostadinov *et al.* (2023) and the different components of the algorithm are summarised below.

Particle size distribution

The size distribution of particles suspended in near-surface ocean waters can be described as a power law, given in differential form as (e.g., Bader, 1970; Sheldon *et al.*, 1972; Jonasz, 1983; Boss *et al.*, 2001; Twardowski *et al.*, 2001; Kostadinov *et al.*, 2009; Roy *et al.*, 2017):

$$\frac{\mathrm{d}N_T}{\mathrm{d}D} = N(D) = N_0 \left(\frac{D}{D_0}\right)^{-\xi}, \qquad (2.2.1)$$

where D is the particle diameter; N (m^{-4}) is the differential number concentration of particles per unit volume seawater and per bin width of particle diameter; N_0 is the particle number concentration at reference diameter D_0, here 2 µm; and ξ is the power-law slope of the PSD. Equation 2.2.1 has to be integrated over a given diameter range to get the total particle number concentration in that range, N_T (m^{-3}).

The PSD algorithm of Kostadinov *et al.* (2023) is mechanistic to the extent feasible, i.e., based on first principles and causality. Some optical model inputs are varied in a Monte Carlo simulation in order to assess uncertainty and base the PSD inversion on an ensemble of forward runs rather than a single set of inputs. Some key simplifying assumptions are made in order to construct an algorithm that is applicable to multispectral ocean colour sensors: 1) Phytoplankton and NAPs have a power-law PSD (i.e., Equation 2.2.1) with the same slope ξ, and 2) The scaling parameter N_0 for NAPs is twice that of N_0 for phytoplankton. The latter assumption is chosen so that it results in a phytoplankton-carbon-to-particulate-organic-carbon ratio of 1 to 3 (see Kostadinov *et al.*, 2016; Behrenfeld *et al.*, 2005), as long as they are both estimated using the same size ranges. Together,

these assumptions allow for the retrieval of one common PSD parameter set pertaining to the total particle population PSD (one ξ value and one total N_0 equal to the linear sum of the phytoplankton and NAPs N_0 values).

Specification of input to particle optical model for phytoplankton and non-algal-particles

The first particle population that contributes to bulk backscattering in the PSD algorithm is represented by living phytoplankton cells, which are modelled using the Aden-Kerker method for coated spheres (Aden and Kerker, 1951). Coated spheres consist of concentric layers/shells of different material properties and can be used to represent phytoplankton cells and their internal heterogeneity and composition (Bernard *et al.*, 2009; Robertson Lain and Bernard, 2018). In the algorithm of Kostadinov *et al.* (2023), the particle coat is represented by the chloroplast and the particle core is represented by the cytoplasm.

To obtain the optical properties of the chloroplast, its relative volume (V_s) is picked from a distribution between 5 and 35% and its imaginary refractive index (relative to seawater) at 675 nm ($n'(675)$) is computed as (Morel and Bricaud, 1986; Bernard *et al.*, 2009; Robertson Lain and Bernard, 2018):

$$n'(675) = \frac{Chl^* \times Chl_i \times 10^6 \times 675 \times 10^{-9}}{4\pi \times V_s \times n_{sw}(675)}, \quad (2.2.2)$$

where Chl^* = 0.027 m^2 mg^{-1} is the theoretical maximum specific absorption coefficient of chlorophyll-a at 675 nm when dissolved in water (Bernard *et al.*, 2009; Robertson Lain and Bernard, 2018); Chl_i is the intracellular chlorophyll-a concentration in kg m^{-3} of cellular material; and $n_{sw}(675)$ is the absolute real refractive index of seawater at 675 nm. A hyperspectral basis vector from the EAP model (based on measurements; see Bernard et al., 2009; Robertson Lain and Bernard, 2018) is then scaled using the value at 675 from Equation 2.2.2, obtaining a hyperspectral relative imaginary refractive index for the chloroplast as the particle coat. In Equation 2.2.2, Chl_i applies to the whole cell and is therefore scaled using V_s to obtain $n'(675)$ for the particle coat alone. The nominal chloroplast's relative real refractive index is then picked from a distribution between 1.06 and 1.22 and modified as a function of its imaginary refractive index according to the Kramers–Kronig relations (implemented as a Hilbert transform; see Bernard *et al.*, 2009; Robertson Lain and Bernard, 2018).

To obtain the optical properties of the cell cytoplasm, its relative real refractive index is picked from a distribution between 1.01 and 1.03, and it is modified by the Kramers–Kronig relations using a constant hyperspectral detritus-like imaginary refractive index, i.e., having an exponential spectral shape similar to that of coloured dissolved organic matter, resulting in a spectrally varying hyperspectral relative real refractive index of the particle core. Details of how each input parameter for phytoplankton cells is specified and the statistical distributions for the Monte Carlo simulations are provided in Tables 1 and 2 of Kostadinov *et al.* (2023).

The second particle population that contributes to bulk backscattering in the PSD algorithm is represented by NAPs, which are modelled using Mie theory (Mie, 1908) for homogeneous spherical particles. For NAPs, the relative refractive indices are picked so that its absorption spectrum is detritus-like (same as the core of phytoplankton), and its real refractive index is allowed to vary over a wider range of values, meant to represent mostly organic detritus, but with some minerogenic contributions, resulting in a mean nominal relative real refractive index of ~1.06. Details of how each input parameter for NAPs is specified and the statistical distributions for the Monte Carlo simulations are provided in Tables 1 and 2 of Kostadinov *et al.* (2023).

Backscattering calculations

The inputs to the particle optical model for phytoplankton and NAPs described above were used to compute the backscattering efficiencies ($Q_{bb}(\lambda)$) for a single phytoplankton cell and a single NAP following Zhang *et al.* (2002) for both coated and homogeneous spheres. Calculations were run for 3,000 instances of Monte Carlo simulations, each with a unique randomly picked combination of

inputs for phytoplankton and NAPs, resulting in 3,000 sets of hyperspectral Q_{bb} values. High sampling resolution in D was used for the coated spheres (10,000 samples between minimum and maximum D) in order to minimise the influence of resonance spikes in Q_{bb}. For NAPs, 1,000 samples of D were used. The computations were performed from 400 to 700 nm wavelength *in vacuo* with a step of 1 nm, allowing the resulting hyperspectral Q_{bb} values to be adapted for use with any combination of visible optical wavebands of multispectral or hyperspectral ocean colour sensors.

Before the calculation of particle backscattering (b_{bp}), the Q_{bb} for each Monte Carlo run were first preprocessed by applying quality control and band-averaging using a moving-average 11 nm wide top-hat filter (using as central wavelengths the nominal bands of SeaWiFS, MERIS, MODIS-Aqua, VIIRS-SNPP and OLCI, plus 440 and 550 nm), resulting in 19 unique bands for band-averaged backscattering efficiencies ($Q_{bb}(\bar{\lambda})$). The band-averaged spectral particulate backscattering coefficient ($b_{bp}(\bar{\lambda})$) was then calculated from $Q_{bb}(\bar{\lambda})$ and the input PSD as follows (Van de Hulst, 1981; Kostadinov *et al.*, 2009):

$$b_{bp}(\bar{\lambda}) = \int_{D_{min}}^{D_{max}} \frac{\pi}{4} D^2 Q_{bb}(D, \bar{\lambda}, m) N_0 \left(\frac{D}{D_0}\right)^{-\xi} \mathrm{d}D \ , \qquad (2.2.3)$$

where m is the complex index of refraction (specified separately for coat and core in the case of phytoplankton). Equation 2.2.3 is applied separately to the modelled phytoplankton and NAP Q_{bb} values and for each of the 3,000 Monte Carlo runs. Band-averaged total $b_{bp}(\bar{\lambda})$ spectra are then calculated as the linear sum of phytoplankton and NAP backscattering.

PSD retrieval via spectral angle mapping

Band-averaged total $b_{bp}(\bar{\lambda})$ spectra were then used to construct the backscattering end-members ($E(\bar{\lambda})$) corresponding to specific input values of the PSD slope (i.e., ξ). First, individual total b_{bp} spectra from each Monte Carlo run were normalised by the value at 555 nm. The median of all normalised spectra at each waveband was used as the end-member for each PSD slope, from ξ = 2.5 to ξ = 6 in steps of 0.05 (see Table 1 in Kostadinov *et al.*, 2023). This approach allowed the isolation of the b_{bp} spectral shape (dependent on ξ) and spectral magnitude (dependent on N_0) (see Equation 2.2.3). Using the hyperspectral underlying Q_{bb} values, end-members can be constructed for any desired set of wavelengths.

The PSD parameters ξ and N_0 are retrieved using $E(\bar{\lambda})$ via the spectral angle mapping (SAM) technique (Dennison *et al.*, 2004). Briefly, the end-members and satellite-observed b_{bp} spectra are treated as n-dimensional vectors, where n is the number of bands. The spectral angle (Θ) between a given end-member and the observed spectrum is then calculated using the vector dot product as:

$$\Theta = \cos^{-1}\left(\frac{b_{bp}(\bar{\lambda}) E(\bar{\lambda})}{\|b_{bp}(\bar{\lambda})\| \|E(\bar{\lambda})\|}\right) . \qquad (2.2.4)$$

Thus, the Θ is an index of spectral shape similarity between two spectra, with more similar spectral shapes resulting in lower Θ. Equation 2.2.4 was used to calculate the Θ between each of the 71 $E(\bar{\lambda})$ and the input observed $b_{bp}(\bar{\lambda})$ spectrum. The value of ξ corresponding to the smallest Θ is then assigned as the retrieved PSD slope. Three wavebands were used, namely 490, 510 and 550 nm. For operational application to OC-CCI v6.0 remote-sensing reflectance ($R_{rs}(\lambda)$) data (which do not have the 550 nm band), band-shifting was applied to the input $R_{rs}(560)$ to estimate the corresponding $R_{rs}(550)$, which is used in the inherent optical properties inversion model of Loisel and Stramski (2000). The band-shifting was constructed using the band ratios between the respective original and target bands from a hyperspectral run of the Morel and Maritorena (2001) model. No other bands were shifted.

The N_0 parameter is subsequently retrieved as the ratio of 1) the satellite-observed value of $b_{bp}(433)/N_0$ and 2) the median value of the quantity $b_{bp}(433)/N_0$ corresponding to the end-member class of the retrieved ξ and all statistically similar classes across all Monte Carlo simulations.

Deriving size-partitioned phytoplankton carbon

Once the PSD parameters are known, they can be used to compute phytoplankton carbon (C in mg m^{-3}) in any size class spanning from cell diameter D_{min} to cell diameter D_{max} as:

$$C = \int_{D_{min_\phi}}^{D_{max_\phi}} 10^{-9} a \left(10^{18} \frac{\pi}{6} D^3\right)^b N_{0_\phi} \left(\frac{D}{D_0}\right)^{-\xi} \mathrm{d}D \ , \quad (2.2.5)$$

where N_{0_ϕ} is $\frac{1}{3}N_0$, and N_0 (m^{-4}) for the total PSD is the satellite-retrieved parameter from total particulate backscattering; the other PSD parameters are as in Equation 2.2.1. The allometric coefficients of Roy *et al.* (2017) are used here, namely a = 0.54 and b = 0.85. When cell volume V is expressed in μm^3, cellular carbon is computed in pg C cell^{-1} using these coefficients and therefore conversion factors were used to convert from m^3 to μm^3 and from pg to mg of carbon (Kostadinov *et al.*, 2016; Roy *et al.*, 2017). Equation 2.2.5 can be used to compute phytoplankton carbon concentrations (mg m^{-3}) in three size classes, namely picophytoplankton (0.2-2 μm), nanophytoplankton (2-20 μm) and microphytoplankton (20-50 μm), as well as total phytoplankton carbon concentration as the sum of the three size classes.

Algorithm tuning

Kostadinov *et al.* (2023) found that N_0 is underestimated at low values and overestimated at high values in a direct comparison with *in situ* observations and a separate cluster was associated with low chlorophyll-a concentrations. Since N_0 is the PSD scaling parameter, which generally controls absolute number, volume and carbon concentration variability to the first order, this could have implications for the global pattern of phytoplankton carbon retrievals and would be consistent with underestimation in the oligotrophic gyres and overestimation in the eutrophic areas. Kostadinov *et al.* (2023) therefore suggest linear tuning of N_0 (in log10 space) for better retrieval of the derived products, following:

$$N_{0_{\text{tuned}}} = 10^{0.3859 \log 10(N_0) + 9.5531} \ , \quad (7)$$

where N_0 is the original (untuned) PSD parameter. This tuning changes N_0 retrievals in a similar fashion to the tuning of the algorithm of Kostadinov *et al.* (2016) and is consistent with a tuning suggested by the N_0 *in situ* validation, i.e., namely, low satellite N_0 values are increased, and high N_0 values are decreased, narrowing the overall range of variability of retrieved N_0 and thus the range of the retrieved derived variables. This addresses the low bias in oligotrophic gyres and the high bias in eutrophic areas. The goal of the tuning is to get more realistic absolute retrievals of particulate organic carbon and chlorophyll-a, and it is suggested that this should also lead to more realistic retrievals of phytoplankton carbon.

Output data

The satellite-based phytoplankton carbon model of Kostadinov *et al.* (2023) estimates the phytoplankton carbon concentration (mg C m^{-3}) for the pico-, nano-, and microphytoplankton size classes and the total phytoplankton community at 4 km spatial resolution for each month between 1997 and 2025 (see an example of May 2015 in Figure 3).

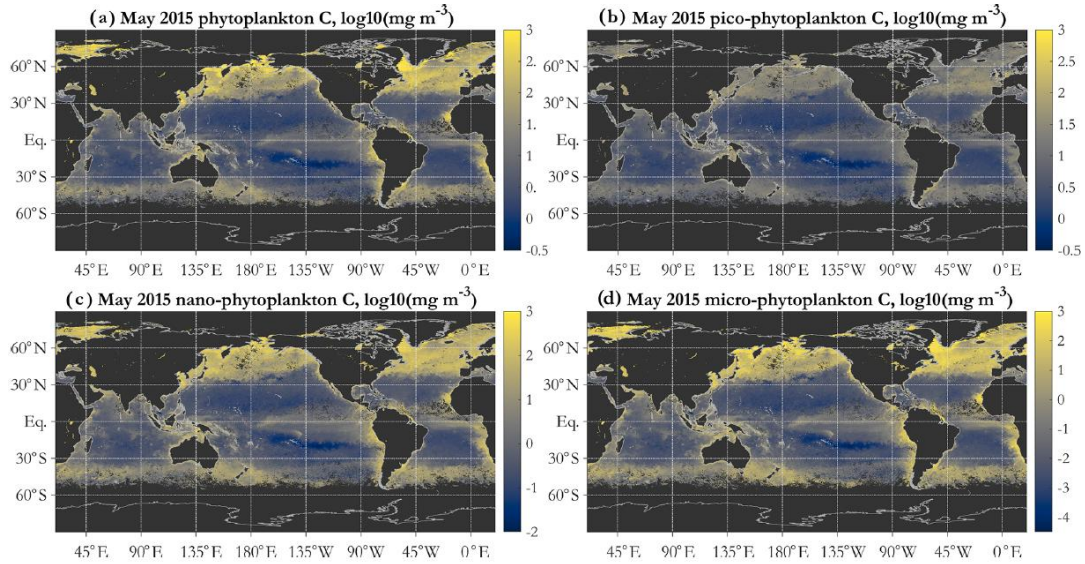

Figure 3. Example of phytoplankton carbon (C) concentration for a) the total phytoplankton community and the b) picophytoplankton, c) nanophytoplankton and d) microphytoplankton size classes in May 2015 for algorithm B. Figure from Kostadinov et al. *(2023).*

2.2.3 Quality assessment

Kostadinov *et al.* (2023) validated their algorithm using *in situ* observations of PSD, particulate organic carbon and the carbon concentration of picophytoplankton as well as concurrent satellite observations of chlorophyll-a, noting that *in situ* data are generally sparce for these variables. The relationship between *in situ* and satellite observations of picophytoplankton carbon showed statistically significant regression, albeit noisy with a relatively low r^2 (Figure 4). For the untuned algorithm, *in situ* and satellite observations covered approximately the same range, and increasing *in situ* picophytoplankton carbon generally corresponded to increasing satellite values as well, with some tendency for under- and overestimation. However, the tuned algorithm showed values of picophytoplankton carbon with narrower range that does not cover the range of the *in situ* observations, and validation statistics are generally worse than those of the untuned algorithm (i.e., the regression is not significant at the $p < 0.05$ level). Kostadinov *et al.* (2023) reported that these validation results were generally consistent with the results of Martínez-Vicente *et al.* (2017), for which the tuned version of the earlier algorithm by Kostadinov *et al.* (2016) was used.

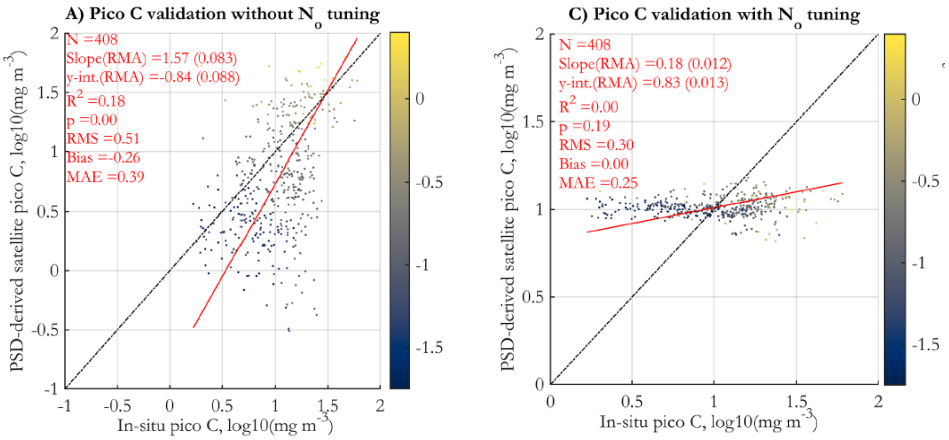

Figure 4. Comparison of satellite-derived picophytoplankton carbon (C) concentration with in situ *data for the A) tuned and C) untuned algorithm B. Figure adjusted from Kostadinov* et al. *(2023). Abbreviations: N = number of observations, RMA = Reduced Major Axis, R2 = Coefficient of determination, p = p-value, RMS = Root-Mean-Square difference, MAE = Mean Absolute Error.*

2.3 Algorithm C

2.3.1 Description

The third phytoplankton carbon algorithm is a chlorophyll-a-based empirical algorithm based on the work of Marañón *et al.* (2014). Chlorophyll-a-based empirical algorithms of phytoplankton carbon use chlorophyll-a as an unequivocal measure for phytoplankton concentration. The method of Sathyendranath *et al.* (2009) was the first of such algorithms that was applied to ocean colour remote sensing observations. It was based on an analysis of Particulate Organic Carbon (POC) as a function of chlorophyll-a and is therefore designed to provide an upper limit on phytoplankton carbon concentrations. Marañón *et al.* (2014) established a relationship between the concentration of chlorophyll-a and carbon in phytoplankton assemblages collected from various oceanic regimes, including temperate, oligotrophic, upwelling and coastal regions, representing a broad range of environmental conditions (Figure 5). This relationship between chlorophyll-a and phytoplankton carbon was shown to be robust, regardless of the potential uncertainties associated with the *in situ* measurements and methods used for conversion of data. While the algorithm of Marañón *et al.* (2014) is based on *in situ* observations, it has since been used for application with remote sensing data (Maniaci *et al.*, 2022) and assessed in intercomparison studies for phytoplankton carbon satellite retrieval algorithms (Martinez-Vicente *et al.*, 2017)

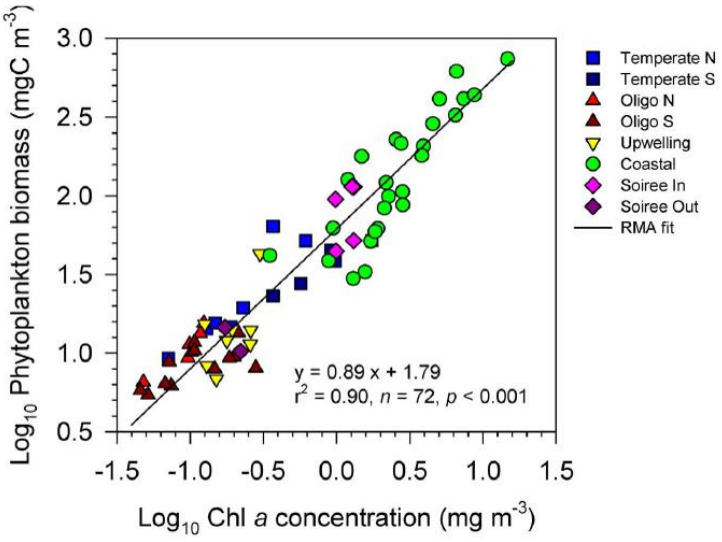

Figure 5. The relationship between chlorophyll-a and phytoplankton carbon from Marañón et al. (2014). Figure adapted from Marañón et al. (2014).

2.3.2 Algorithm definition

Input data

The following input data are required for the satellite-based phytoplankton carbon algorithm:

- Chlorophyll-a concentration from OC-CCI version 6.0

All data are required at 4 km spatial and monthly temporal resolution.

Retrieval algorithm

The algorithm of Marañón *et al.* (2014) is an empirical algorithm that describes the relationship between chlorophyll-a and phytoplankton carbon biomass, based on concurrently collected *in situ*

observations in surface assemblages of coastal and open ocean regions. Phytoplankton carbon (C in mg m^{-3}) can be estimated as:

$$C = 10^{0.89 \times \log(B) + 1.79}, \quad (2.3.1)$$

where B is the chlorophyll-a concentration (in mg m^{-3}).

Output data

The satellite-based phytoplankton carbon model estimates the phytoplankton carbon concentration (mg C m^{-3}) for the total phytoplankton community at 4 km spatial resolution for each month between 1997 and 2025 (see an example of the climatology in Figure 6).

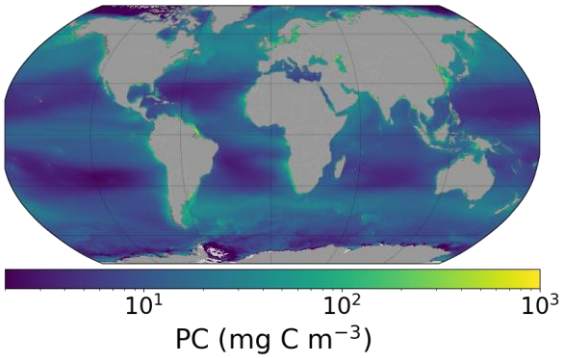

Figure 6. Climatology of phytoplankton carbon (PC) concentration of the total phytoplankton community for 2000-2020 for algorithm C.

2.3.3 Quality assessment

Marañón *et al.* (2014) reported that the observed relationship between chlorophyll-a and phytoplankton carbon was remarkably robust, yet that the slope of the log-log regression was significantly lower than 1, indicating that the carbon-to-chlorophyll ratio tends to decrease as phytoplankton standing stocks increase from open-ocean oligotrophic to open-ocean temperate and then coastal waters. This pattern likely reflects the fact that in temperate and coastal waters phytoplankton are acclimated to lower mean irradiances, which results in higher cellular concentrations of chlorophyll-a (Marañón *et al.*, 2014). In an intercomparison of satellite-retrieval algorithms of picophytoplankton carbon (i.e., phytoplankton with a cell size <2 µm), chlorophyll-a-based empirical algorithms performed well compared with other algorithms, especially at chlorophyll-a concentrations less than 0.15 mg m^{-3} (Martinez-Vicente *et al.*, 2017). The algorithm of Marañón *et al.* (2014), applied to chlorophyll-a data from version 2 of OC-CCI, showed a bias of -0.02 and Root Mean Square Difference (RMSD) of 0.33 compared with *in situ* observations of picophytoplankton carbon derived from flow cytometry measurements (bias of all algorithms ranged between -0.24 and +0.33 and RMSD of all algorithms ranged between 0.33 and 0.57; Martinez-Vicente *et al.*, 2017).

2.4 Algorithm D

2.4.1 Description

The fourth satellite-based phytoplankton carbon model is based on the coupled photoacclimation and primary production model of Sathyendranath *et al.* (2020). The primary production model is based on the concentration of chlorophyll-a, which is readily obtained from satellite data. To convert the carbon fixed through primary production to the corresponding increase in chlorophyll-a units,

information on the chlorophyll-to-carbon ratio is needed, which is highly variable and depends on the photoacclimation status of phytoplankton. With the formulation of a resource allocation model that is based on the same set of parameters used in the primary production model (Jackson *et al.*, 2017b), Sathyendranath *et al.* (2020) established the maximum chlorophyll-to-carbon ratio that phytoplankton can attain when the available light tends to zero, information needed to implement the coupled model for the estimation of phytoplankton carbon. With the maximum chlorophyll-to-carbon ratio established for different phytoplankton size classes, the coupled model can provide information on the carbon content of pico-, nano- and microphytoplankton as well as the carbon content of the total phytoplankton community.

2.4.2 Algorithm definition

Input data

The following input data are required for the satellite-based phytoplankton carbon algorithm:

- Chlorophyll-a concentration from OC-CCI version 6.0
- Photosynthetically Active Radiation (PAR) from NASA, a merged sensor-product
- Maps of photosynthesis versus irradiance parameters
- Maps of chlorophyll-a profile parameters
- Bathymetry from GEBCO
- Mixed layer depth from De Boyer Montégut *et al.* (2004)

All data are required at 4 km spatial and monthly temporal resolution.

Retrieval algorithm

The theoretical considerations that underpin the coupled photoacclimation and primary production model are provided below, to understand how phytoplankton photoacclimation and photosynthesis change relative to each other, which is essential information to derive the chlorophyll-to-carbon ratio. The phytoplankton carbon model is based on a photoacclimation model that uses the same photosynthetic parameters as the coupled primary production model (Platt and Sathyendranath, 1988; updated as in Sathyendranath *et al.*, 2020 and Kulk *et al.* 2020, 2021), i.e., the initial slope of the photosynthesis versus irradiance (P-I) curve α^B (in mg C mg Chl-a^{-1} h^{-1} (µmol photons m^{-2} s^{-1})$^{-1}$) and the assimilation number of the P-I curve P_m^B (in mg C mg Chl-a^{-1} h^{-1}). The ratio of these two parameters is known as the photoacclimation parameter I_k (in µmol photons m^{-2} s^{-1}):

$$I_k = P_m^B/\alpha^B. \quad (2.4.1)$$

When the ambient irradiance I (i.e., PAR in µmol photons m^{-2} s^{-1}) is normalised to I_k, the dimensionless irradiance I_* can be defined as (also see Platt and Sathyendranath, 1993):

$$I_* = I/I_k. \quad (2.4.2)$$

With these photosynthetic parameters, and the definition of I_k, we can write the light-saturation curve as (Platt *et al.*, 1980):

$$P^B(I_*) = P_m^B(1 - \exp(-I_*)), \quad (2.4.3)$$

where P^B is primary production normalised to the chlorophyll-a concentration (in mg C mg Chl-a^{-1} h^{-1}).

Regarding the chlorophyll-to-carbon ratio (θ, in mg Chl mg C^{-1}), an analytical solution to the model of Geider *et al.* (1997) was published by Jackson *et al.* (2017b):

$$\theta = \left(\frac{\theta_m}{I_*}\right)(1 - \exp(-I_*)) \,, \tag{2.4.4}$$

where θ_m is the maximum attainable value of θ (a constant of particular phytoplankton species or types, to be determined) (Geider *et al.,* 1997). Comparing the right-hand sides of Equations 2.4.3 and 2.4.4, we find:

$$\left(\frac{1}{I_*}\right)\left(\frac{P^B}{P_m^B}\right) = \frac{\theta}{\theta_m} \,; \tag{2.4.5}$$

and in the special case where I_* = 1:

$$\frac{P^B}{P_m^B} = \frac{P}{P_m} = \frac{\theta}{\theta_m} \,. \tag{2.4.6}$$

Here P is absolute production (in mg C h^{-1}, i.e., not normalised to the chlorophyll-a concentration) and P_m is the maximum production (in mg C h^{-1}, also unnormalised). Equation 2.4.6, which is independent of the chlorophyll-a concentration, represents the most economical way to express the reconciliation of primary production models with models of the chlorophyll-to-carbon ratio (Figure 7).

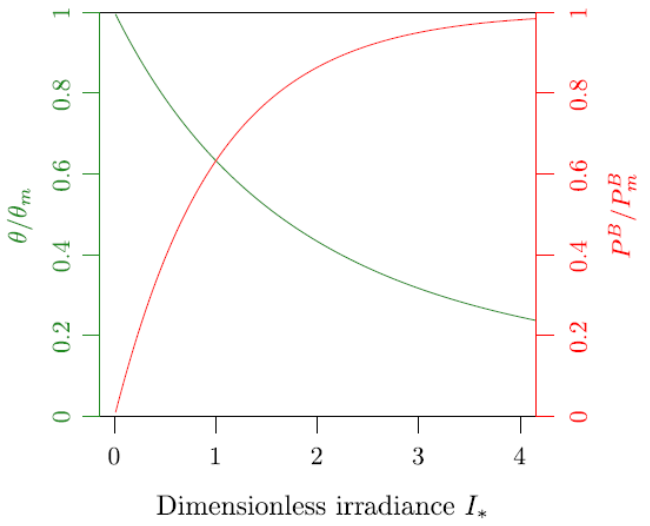

Figure 7. Curves of primary production, normalised to its maximum value P_m^B at saturating light, and θ normalised to its maximum value θ_m at zero irradiance, as functions of the dimensionless irradiance I_*. The two curves coincide when the dimensionless irradiance is 1, demonstrating the connection between the two models.

With this theoretical basis, and under the assumption of balanced growth (Geider *et al.*, 1996, 1997), the changes of carbon (C) and chlorophyll (B) through time in the surface layer of the ocean can be explored. Let us consider production ($P_{Z_m,T}$) in the layer extending from the surface to the mixed-layer depth (Z_m in m) and through the daylight hours (D in h), with $\overline{P_{Z_m,T}}$ as the average production over depth and time. Assuming that within the mixed layer, the chlorophyll-a concentration and model parameters are uniform with depth, and considering two consecutive days, 0 and 1, the carbon budget for the mixed layer will be:

$$C_1 - C_0 = \overline{P_{Z_m,T}} - C_r \,, \tag{2.4.7}$$

where C_r represents carbon removed by various loss processes, including respiration, export, and grazing. The corresponding chlorophyll-a budget is:

$$B_1 - B_0 = \theta(C_1 - C_0) \ . \tag{2.4.8}$$

Comparing the two budgets in Equations 2.4.7 and 2.4.8, we find:

$$\frac{B_1 - B_0}{\overline{P_{Z_m,T}} - C_r} = \theta \ . \tag{2.4.9}$$

In Equation 2.4.9, all quantities on the left-hand side are known, with the exception of C_r. In the special case that C_r is set to zero, we have:

$$\theta = \frac{B_1 - B_0}{\overline{P_{Z_m,T}}} \ , \tag{2.4.10}$$

which provides a lower bound on θ.

Turning to the calculation of $P_{Z_m,T}$, we have to integrate the light-saturation curve (i.e., Equation 2.4.3) through the mixed layer and over the day:

$$P_{Z_m,T} = B \int_0^D \int_0^{Z_m} P^B(z,t) \mathrm{d}z \mathrm{d}t \ , \tag{2.4.11}$$

where D is the day length (in h). Substituting for the depth and time dependence of the irradiance, as well as for the functional form of the light-saturation curve and converting to a spectral form by noting that the wavelength-dependent terms, α and I, will always occur as a product of the form αI in all primary production models (Sathyendranath and Platt, 1989, 2007), we have:

$$P_{Z_m,T} = B P_m^B \int_0^D \int_0^{Z_m} (1 - \exp(-\Pi(\mathrm{z},\mathrm{t})/P_m^B)) \mathrm{d}z \mathrm{d}t \ , \text{ where} \tag{2.4.12}$$

$$\Pi(\mathrm{z},\mathrm{t}) = \int \alpha(\lambda,\mathrm{z},\mathrm{t}) I(\lambda,\mathrm{z},\mathrm{t}) d\lambda \ . \tag{2.4.13}$$

Note that these equations are the unnormalised forms of the equations presented for the primary production model of Platt and Sathyendranath (1988) as updated by Sathyendranath *et al.* (2020) and Kulk *et al.* (2020, 2021).

The corresponding dimensionless irradiance can be written as:

$$I_*(z,t) = \Pi(\mathrm{z},\mathrm{t})/P_m^B \ , \tag{2.4.14}$$

and its average over the day length and over the mixed layer is then given by:

$$\langle I_* \rangle_{Z_m T} = \frac{1}{D Z_m} \int_0^D \int_0^{Z_m} I_*(z,t) \mathrm{d}z \mathrm{d}t \ . \tag{2.4.15}$$

As is the case for production, the subscripts Z_m and T indicate averaging with respect to the mixed layer depth and time of day (from dawn to dusk).

The theory described above was implemented by Sathyendranath *et al.* (2020) using satellite data to compute primary production and mean I_* (over the day length and for the mixed layer) for a sample image of the Ocean Colour Climate Change Initiative (OC-CC) for two randomly selected consecutive days (1 and 2 May 2010) to estimate the upper bound of θ_m for each of the ecological provinces as defined by Longhurst (2007). This approach led to values of θ_m ranging between 0.0033 and 0.081 (mean = 0.024 and standard deviation = 0.017).

The established range of θ_m was used to estimate the phytoplankton carbon content for three phytoplankton size classes: picophytoplankton (<2 μm), nanophytoplankton (2-20 μm) and

microphytoplankton (>20 µm). First, phytoplankton biomass is estimated for pico-, nano- and microphytoplankton using the three-component model of Brewin *et al.* (2011, 2015a):

$$B_p = B_p^m\left[1 - \exp\left(-S_p B\right)\right] , \quad (2.4.16)$$

$$B_{p,n} = B_{p,n}^m\left[\left(1 - \exp\left(-S_{p,n} C\right)\right]\right], \quad (2.4.17)$$

$$B_n = B_{p,n} - B_p \text{ , and} \quad (2.4.18)$$

$$B_m = B - B_{p,n} . \quad (2.4.19)$$

Here, the fitted parameters B_p^m (0.13 mg Chl-a m^{-3}) and $B_{p,n}^m$ (0.77 mg Chl-a m^{-3}) are the maximum chlorophyll-a concentrations attainable by picophytoplankton and combined pico- and nanophytoplankton, respectively. The parameters S_p (6.15 m^3 mg Chl-a^{-1}) and $S_{p,n}$ (1.26 m^3 mg Chl-a^{-1}) determine the rate of change in the chlorophyll-a concentrations associated with picophytoplankton and the combined concentration of pico- and nanophytoplankton with changes in total chlorophyll-a concentration (model parameters are from Brewin *et al.* (2015a). The phytoplankton carbon concentration of the three different phytoplankton size classes (C_p, C_n and C_m in mg C m^{-3}) can then be calculated using their respective chlorophyll-a concentration (i.e., Equations 2.4.16 to 2.4.19) and the mean I_* over the mixed layer (i.e., Equation 2.4.15):

$$C_p = \frac{B_p * \langle I_* \rangle_{Z_m,T}}{\theta_{max}\left(1 - e^{-\langle I_* \rangle_{Z_m,T}}\right)} , \quad (2.4.20)$$

$$C_n = \frac{B_n * \langle I_* \rangle_{Z_m,T}}{\theta_{max}\left(1 - e^{-\langle I_* \rangle_{Z_m,T}}\right)} , \quad (2.4.21)$$

$$C_m = \frac{B_m * \langle I_* \rangle_{Z_m,T}}{\theta_{max}\left(1 - e^{-\langle I_* \rangle_{Z_m,T}}\right)} , \quad (2.4.22)$$

where the θ_m is 0.03, 0.06 and 0.08 for pico-, nano- and microphytoplankton, respectively. The total phytoplankton carbon concentration (in mg C m^{-3}) is the sum of that of the three phytoplankton size classes:

$$C = C_p + C_n + C_m . \quad (2.4.23)$$

Output data

The satellite-based phytoplankton carbon model estimates the phytoplankton carbon concentration (mg C m^{-3}) for the pico-, nano-, and microphytoplankton size classes and the total phytoplankton community at 4 km spatial resolution for each month between 1997 and 2025 (see an example of the climatology in Figure 8).

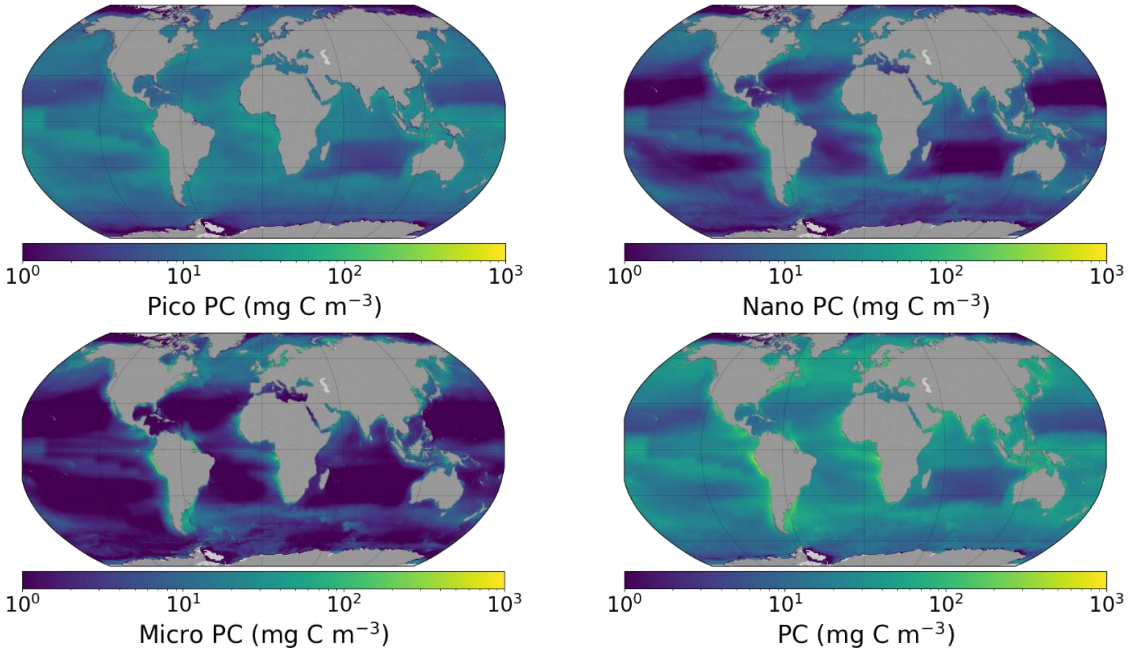

Figure 8. Climatologies of phytoplankton carbon (PC) concentration for pico-, nano-, and microphytoplankton size classes and the total phytoplankton community for 1998-2025 for algorithm D.

2.4.3 Quality assessment

Comparison of the coupled photoacclimation and primary production model with *in situ* data showed consistency between the satellite-based estimates and *in situ* data, especially considering that the chlorophyll-a ranges for the *in situ* data overlap little with those of the satellite data (Figure 9).

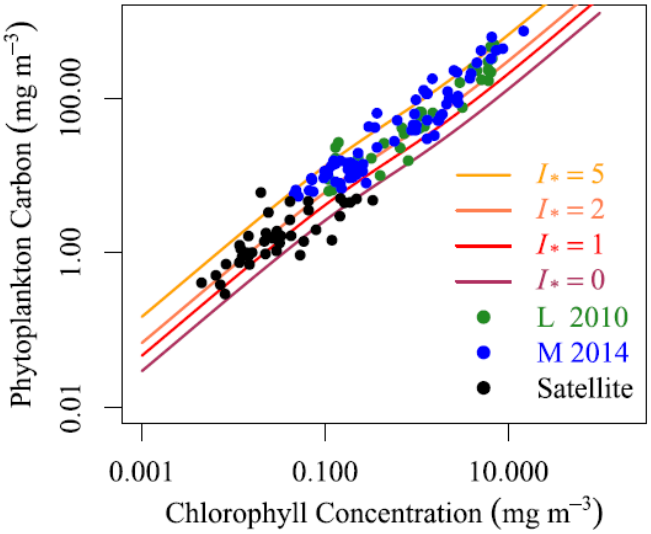

Figure 9. Field data on phytoplankton carbon (mg m^{-3}) and chlorophyll-a concentration (mg m^{-3}) from Li et al. *(2010) (in green) and from Marañón* et al. *2014 (in blue). Also shown are the points (in black) corresponding to the satellite-based estimates of θ_m for different ecological provinces. The curves correspond to model estimates, for I_* = 0, 1, 2, and 5. The curve for I_* = 0 represents θ_m variability with size class.*

3 Phytoplankton pigment diversity

All phytoplankton pigment algorithms considered in PHYTO-CCI were based on existing satellite-retrieval algorithms, but trained using the *in situ* pigment data matched with remote sensing reflectances (R_{RS}) at six wavebands of the Ocean Colour Climate Change Iniative (OC-CCI; version 6) and for some algorithms with sea surface temperature (SST) from SST-CCI (version 3). PHYTO-CCI focuses on the retrieval of eight different phytoplankton pigments (see Section 3.1.2), in addition to chlorophyll-a, which is not further discussed as this variable is already provided by OC-CCI. The *in situ* – satellite match-up dataset used for training of the algorithm are described in the Product Validation and Algorithm Selection Report (PVASR). Algorithms A_1 and B_1 have been extended by incorporating SST, and these algorithms are named respectively, A_2 and B_2. We note that each phytoplankton algorithm described in the next sections is actually a combination of algorithms/models that estimate individual pigments following the same method.

3.1 Algorithm A_1

3.1.1 Description

Algorithm A_1 is based on the use of the spectral signature of satellite observations to predict various phytoplankton pigment concentrations. This spectral algorithm was first proposed by Bracher *et al.* (2015) and further updated by Xi *et al.* (2020) using empirical orthogonal function (EOF) analysis on R_{RS}-based regression to matched *in situ* phytoplankton pigment observations. The original algorithm by Bracher *et al.* (2015) was developed for the Atlantic Ocean and made use of eight bands for multi- or hyperspectral (400 to 550 nm) *in situ* observations and eight R_{RS} bands from MERIS. In Xi *et al.* (2020), the algorithm was extended for global application based on a global match-up dataset of *in situ* phytoplankton pigments to nine spectral bands of the merged R_{RS} dataset of GlobColour (for more information see http://www.globcolour.info/) and 11 R_{RS} bands of Sentinel-3A OLCI. The *in situ* pigments were converted into phytoplankton groups (in units of chlorophyll-a) using the diagnostic pigment analysis by Vidussi *et al.* (2001) as adapted by Uitz *et al.* (2006) and Hirata *et al.* (2011). From these upscaled datasets, finally the chlorophyll-a concentration of Phytoplankton Functional Types (PFTs) were developed. The robustness of the algorithms was verified in both studies (Bracher *et al.*, 2015; Xi *et al.*, 2020) via cross validation. The global PFT chlorophyll-a retrievals of Xi *et al.* (2020) were further evaluated via intercomparison to other PFT retrievals, PHYSAT (Alvain *et al.,* 2005, 2008) and SynSenPFT (Losa *et al.*, 2017). The algorithm (named PFT-EOF) was implemented in 2020 into the Copernicus Marine Service (CMEMS, see https://marine.copernicus.eu/). For PHYTO-CCI, the algorithms by Bracher *et al.* (2015) and Xi *et al.* (2020) were adapted to a global algorithm (named Pig-EOF, i.e., Algorithm A_1) using the six R_{RS} bands from OC-CCI as input data. The global models were developed based on match-ups of these R_{RS} with global *in situ* phytoplankton pigment dataset (see PVASR). Based on the method by Xi *et al.* (2021), pixel-by-pixel uncertainties can also be implemented into the operational processing of these final products. Figure 10 summarises Algorithm A_1 in a flow chart with details of the input (without SST) and output data and the retrieval principle and quality assessment. The algorithm is described in detail in Section 3.1.2.

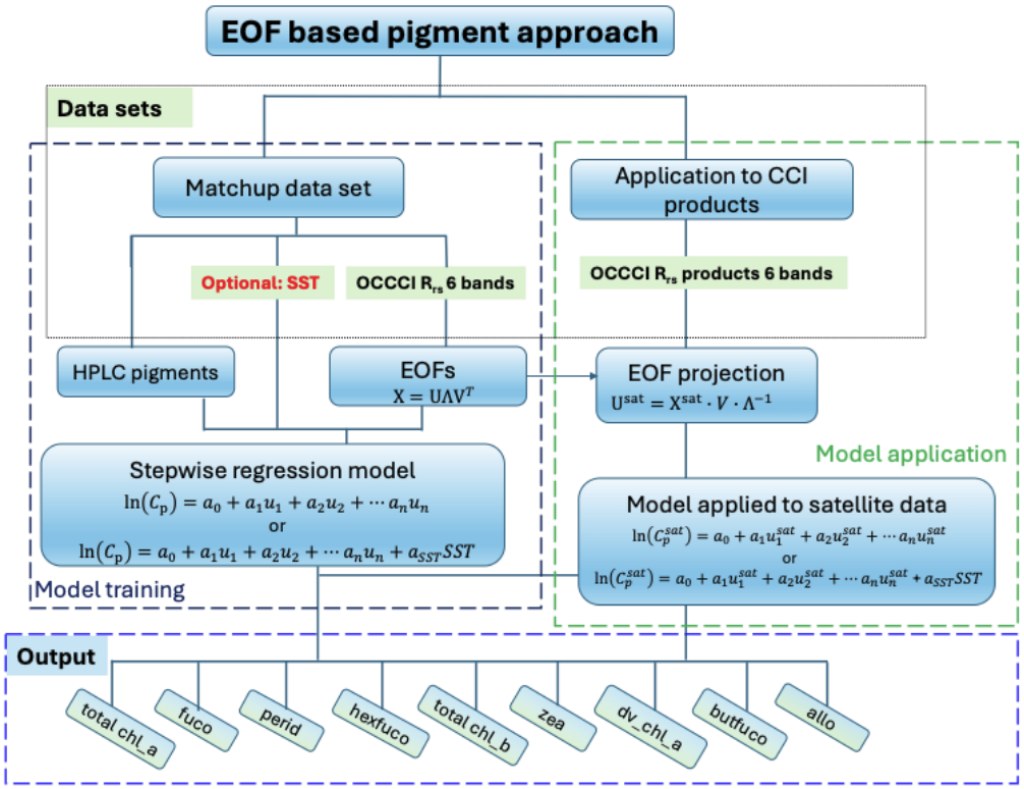

Figure 10. Flowchart of PHYTO-CCI EOF algorithms based on R_{RS} only (Algorithm A_1, i.e., Pig-EOF) and based on R_{RS}+SST (Algorithm A_2, i.e., Pig-EOF-SST).

3.1.2 Algorithm definition

Input data

For algorithm development, match-ups of remote sensing reflectances ($R_{RS}(\lambda)$) at 412, 443, 490, 560, 620 and 665 nm from OC-CCI version 6.0 were used together with *in situ* phytoplankton pigment data derived from High Pressure Liquid Chromatography (HPLC). The *in situ* – satellite match-up dataset used for training of the algorithm is described in detail in the PVASR. An overview of the pigments considered for robust prediction by the algorithm is provided in Table 1.

Table 1. Overview of phytoplankton pigments targeted in PHYTO-CCI.

Pigment	Abbreviation
19'-Butanoylfucoxantin	butfuco
19'-Hexanoylfucoxantin	hexfuco
Alloxanthin	allo
Divinyl chlorophyll-a	dv_chl_a
Total chlorophyll-a	total_chl_a
Total chlorophyll-b	total_chl_b
Fucoxanthin	fuco
Peridinin	peri
Zeaxanthin	zea

After the development of algorithms for each phytoplankton pigment using the *in situ* – match-up dataset, the algorithms can be applied to R_{RS} of OC-CCI to produce global products phytoplankton pigments.

Retrieval algorithm

For each phytoplankton pigment, regression models were developed separately. In brief, matched $R_{RS}(\lambda)$ to each valid pigment concentration from the development dataset were first standardised by subtracting the mean spectral value and then divided by the spectral standard deviation (Taylor *et al.*, 2013; and then used similarly in all related publications using the EOF model, e.g., Bracher *et al.*, 2015). Then the standardised $R_{RS}(\lambda)$ as matrix X (with M observations × N wavelengths), which may vary among different pigments due to the number of match-ups, was subjected to a Singular Value Decomposition (SVD) to derive the EOF scores (U), singular values (Λ) and EOF loadings (V):

$$X = U\Lambda V^T \ . \quad (3.1.1)$$

A generalised linear model (GLM) was built for each pigment, considering the log-transformed *in situ* pigment concentration (from HPLC) and the corresponding EOF scores subset. Insignificant EOF modes were excluded, and a stepwise routine was then adapted to search for the most significant regression model through the minimisation of the Akaike Information Criteria (AIC). The AIC is an estimator of prediction error, which uses the number of scores to determine the best model among the total in a given dataset. Through the AIC, only significant EOF scores were used in the regression model. As a result, the regression model for the pigment concentration prediction was expressed as:

$$\ln(C_{pig}) = a_0 + a_1u_1 + a_2u_2 + \cdots .. a_nu_n \ , \quad (3.1.2)$$

where C_{pig} is the concentration of each pigment (in mg m^{-3}), a_0 is the intercept, $a_{1,2,\ldots,n}$ are the regression coefficients and $u_{1,2,\ldots,n}$ are the number (n) of EOF scores.

Lastly, for application to a new R_{RS} dataset (e.g., satellite products), these EOF-based models were projected to the standardised new data onto the EOF loadings from the training dataset and a new set of EOF scores was derived:

$$U^{new} = X^{new}V\Lambda^{(-1)} \ . \quad (3.1.3)$$

The new specific prediction is then obtained by applying these new EOF scores to Equation 3.1.2.

Output data

The following phytoplankton pigment concentrations (mg m^{-3}) are predicted from the algorithm at 4 km spatial resolution for each month between 1997 and 2025: Alloxanthin, 19'-butanoylfucoxantin, divinyl chlorophyll-a, fucoxanthin, 19'-hexanoylfucoxantin, peridinin, total chlorophyll-a, total chlorophyll-b and zeaxanthin. An example of the monthly products for fucoxanthin, 19'-hexanoylfucoxantin, total chlorophyll and zeaxanthin is shown in Figure 11.

3.1.3 Quality assessment

To assess algorithm performance based on the full development dataset, statistical metrics including the coefficient of determination in log space (Log[R^2]) and the root-mean-square difference (RMSD) and the median absolute percent difference (MdAPD) in linear space were generated. Figure 12 and Table 2 show the full-fit algorithm performance for all pigments, including total chlorophyll-a, but we note that this pigment is not further described as it is not considered within PHYTO-CCI. In addition, a further evaluation of the PHYTO-CCI EOF-algorithm is based on Cross Validation (CV) statistics, following Bracher *et al.* (2015), in order to assess the robustness of the fitted EOF-based pigment concentration regression models by splitting the whole dataset into two subsets; 80% of the data were used for model training and 20% for model validation. This procedure is run for 500 permutations, and during each permutation, EOF predicted and observed pigment concentrations

and the statistical metrics for each permutation were recorded for evaluation. The Log[R^2] of each of the GLM regression, which is based on log-scaled pigment concentration, and then the mean of all permutations (Log[CV.R^2]) was calculated. Similarly, the mean of the RMSD and the MdAPD for all permutations were calculated based on the non-log-transformed pigment concentration data (i.e., CV.RMSD and CV.MdAPD) (for definitions see Bracher *et al.*, 2026).

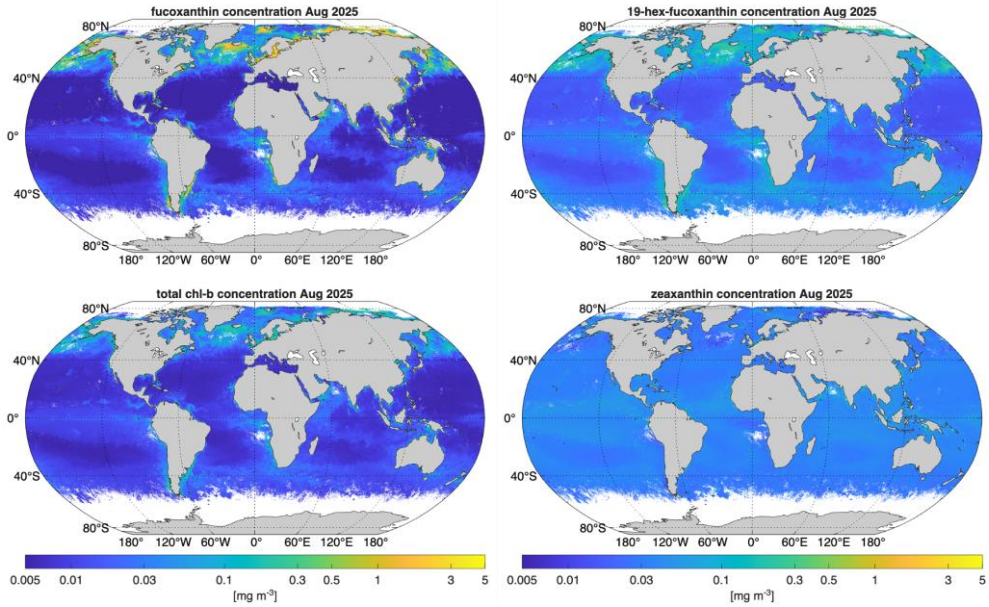

Figure 11. Example (August 2025) of monthly phytoplankton pigment product from Algorithm A_1 (Pig-EOF) for the pigments fucoxanthin, 19'-hexanoylfucoxanthin, total chlorophyll-b (total chl-b) and zeaxanthin.

The results of the full-fit model performance (Figure 12, Table 2) and cross validation statistics (Table 2) of Algorithm A_1 were consistent for all statistical metrics (Log[R^2], MdAPD, RMSD vs. Log[CV.R^2], CV.MdAPD, and CV.RMSD, respectively), which confirmed overall a robustness of the individual pigment models. Agreement of predictions to observations varies across different statistical metrics and specific pigment predictions. Good results in terms of R^2 between 0.5–0.74 were obtained for most pigment models, and only slightly lower for alloxanthin and zeaxanthin (R^2 = 0.45), but much lower for divinyl chlorophyll-a (R^2 = 0.31). Nearly all pigment models achieved acceptable MdAPD ranging from 38% (19'-butanoylfucoxanthin) to 58% (fucoxanthin), only alloxanthin obtained higher values (i.e., 67%). Overall, the results for Algorithm A_1 were slightly better than what was obtained for the first version of PFT-EOF models (for PFT chlorophyll-a concentrations; Xi *et al.*, 2020). Following the results, we decided to select all the PIG-EOF models for intercomparison in PHYTO-CCI. We note that updates of the original PFT-EOF algorithm by Xi *et al.* (2020) implemented in CMEMS showed much better performance than the algorithms reported here. This model was not only achieved by the extension of the match-up dataset for model development, but also by adding additional satellite input variables on which Algorithm A_2 is based on. Overall, performance of Algorithm A_1 in the PHYTO-CCI algorithm comparison (see PVASR) points to large room for improvement and certainly should include as in Algorithm A_2 the addition of SST as an input variable.

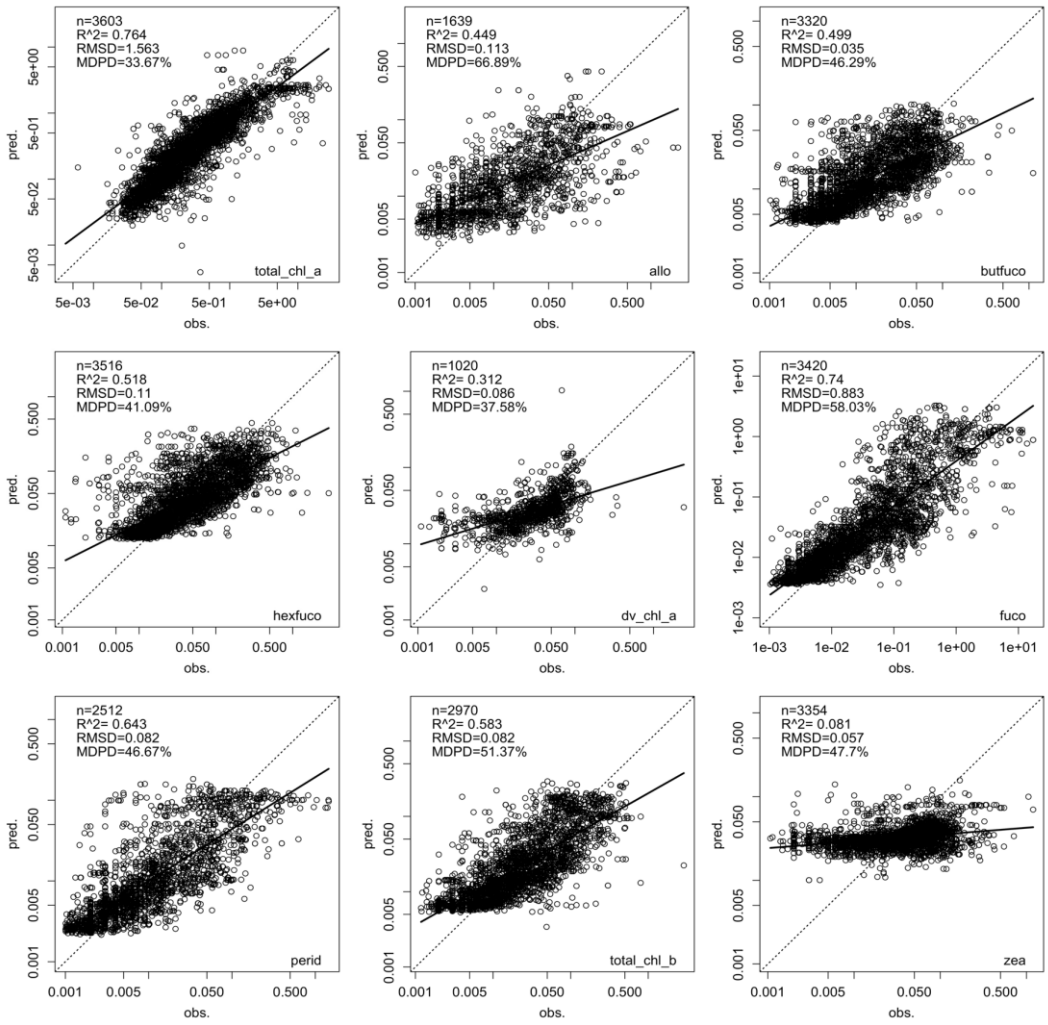

Figure 12. Algorithm A_1 (i.e., Pig-EOF) full-fit model performance statistics and scatterplots of pigment concentration predictions and observation from match-up HPLC pigment data.

Table 2. Full-fit performance and cross validation (CV) statistics of Algorithm A_1. The coefficient of determination (R^2) is based on log-transformed data. The Median Absolute Percentage Difference (MdPAD) is provided as percentage values and the Root Mean Square Difference (RMSD) in absolute units (mg m^{-3}). N is the number of observations.

		Full-fit			Cross validation		
Pigment	**N**	**Log[R^2]**	**RMSD (mg m^{-3})**	**MdAPD (%)**	**Log[CV.R^2]**	**CV.MdAPD (%)**	**CV.RMSD (mg m^{-3})**
butfuco	3320	0.499	0.035	46.287	0.498	46.354	0.034
hexfuco	3516	0.518	0.110	41.095	0.518	41.089	0.110
allo	1639	0.449	0.113	66.887	0.446	67.348	0.108
total_chl_a	3603	0.764	1.563	33.673	0.761	33.899	1.576
dv_chl_a	1020	0.312	0.086	37.581	0.309	37.728	0.065
total_chl_b	2970	0.583	0.082	51.365	0.581	51.675	0.079
fuco	3420	0.740	0.883	58.031	0.737	58.159	0.861
perid	2512	0.643	0.082	46.671	0.641	47.047	0.079
zea	3354	0.081	0.057	47.699	0.079	47.720	0.056

3.2 Algorithm A_2

3.2.1 Description

This algorithm serves as an improved version of algorithm A_1 based on the retuned EOF-based algorithm following Xi *et al.* (2021). The additional variable SST is incorporated into the adapted model, given the fact that most of the pigments (which are the biomarkers of phytoplankton groups) are highly correlated to SST at the global scale. Xi *et al.* (2021), and updates of their model as presented by Xi *et al.* (2023, 2025), demonstrated that incorporating SST leads to much better performance in the EOF-based algorithm for most of the phytoplankton groups, especially for prokaryotic phytoplankton. Figure 10 summarises Algorithm A_2 in a flowchart, detailing the input and output data and the retrieval principle and quality assessment.

3.2.2 Algorithm definition

Input data

In addition to the match-up data of $R_{RS}(\lambda)$ at six bands from OC-CCI version 6.0 and *in* situ concentrations of phytoplankton pigments from HPLC measurements from the algorithm development dataset (see PVASR), SST data is required for development of Algorithm A_2. The SST data were derived from the Sea Surface Temperature Climate Change Initiative (SST-CCI; version 3), as available through the Copernicus Climate Change Service (C3S; https://cds.climate.copernicus.eu/datasets/satellite-ist-sst-global?tab=overview), which includes the climate data record up to 2022 and the interim climate data record up to the present. The same nine phytoplankton pigments listed for Algorithm A_1 (Table 1) were included for predictions.

After the development of the phytoplankton carbon algorithm, the algorithms for each phytoplankton pigment can be applied to $R_{RS}(\lambda)$ data from OC-CCI and SST data from SST-CCI, given that the latter is re-gridded to the same spatial resolution and grid as the OC-CCI data.

Retrieval algorithm

For Algorithm A_2, the EOF analysis remained unchanged, using Singular Value Decomposition (SVD) to decompose the (standardised) R_{RS} spectra into the EOF scores (U), singular values (Λ) and EOF loadings (V) as for Algorithm A_1 (see Equation 3.1.1). When formulating the regression models of pigments, SST is then introduced as an additional term together with the column vectors $u_{1,2,\ldots,n}$ in U. Similar to Algorithm A_1, a stepwise routine was applied to obtain smaller regression models by removing least significant variables in U through minimisation of the Akaike Information Criterion (AIC). The adapted regression model is expressed as:

$$\ln(C_{pig}) = a_0 + a_1u_1 + a_2u_2 + \cdots + a_nu_n + a_{SST}SST \ . \quad (3.2.1)$$

For algorithm implementation, $R_{RS}(\lambda)$ values at the six wavebands of OC-CCI and SST data are prepared at the same spatial scale and the projection on the standardised R_{RS} data is carried out through Equation 3.1.3 as for Algorithm A_1. The predictions of pigments are then obtained by applying the new EOF scores and SST to Equation 3.2.1.

Output data

As for Algorithm A_1 (see Section 3.1.2), Algorithm A_2 predicts nine phytoplankton pigment concentrations (in mg m^{-3}) at 4 km spatial resolution for each month between 1997 and 2025. An example of the monthly products for fucoxanthin, 19'-hexanoylfucoxantin, total chlorophyll and zeaxanthin are shown in Figure 13.

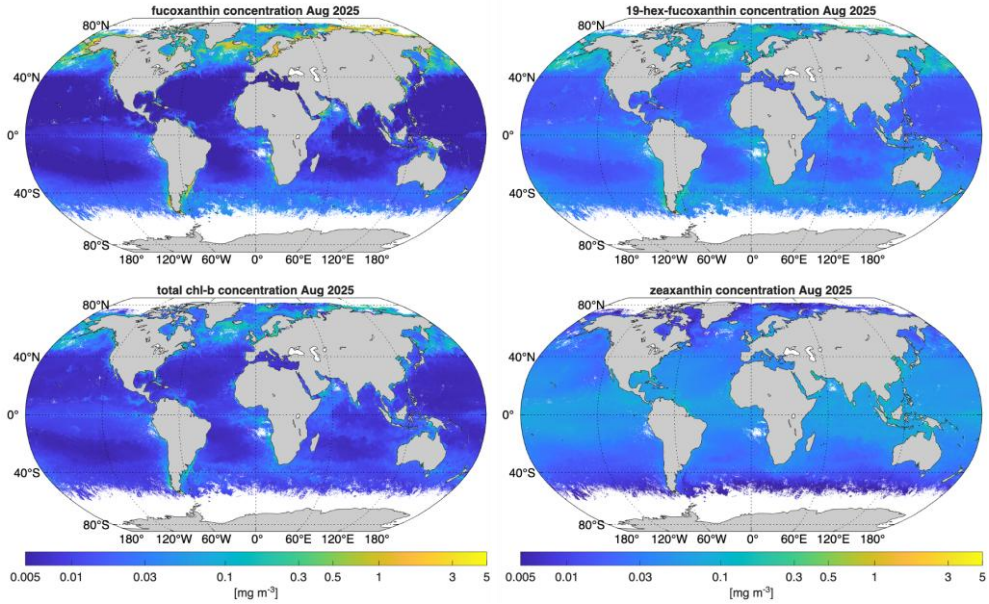

Figure 13. Example (August 2025) of monthly phytoplankton pigment product from Algorithm A_2 (Pig-EOF-SST) for the pigments fucoxanthin, 19'-hexanoylfucoxanthin, total chlorophyll-b (total chl-b) and zeaxanthin.

3.2.3 Quality assessment

The assessment of the Algorithm A_2 followed the exact same procedure as described in Section 3.1.3 for Algorithm A_1. Figure 14 shows the full-fit model performance for all pigments. The results of the model performance and cross-validation statistics of Algorithm A_2 are presented in Table 3. From both sets of statistics, most of the retrievals are slightly better than those of Algorithm A_1, except for 19'-hexanoylfucoxanthin, which remained the same as addition of SST was not considered significant after the AIC criteria. Zeaxanthin showed the highest improvement with SST incorporated (R^2 improved from 0.06 to 0.37 and MdAPD from 48% to 43%). Overall performance of Algorithm A_2 in the intercomparison of PHYTO-CCI (see PVASR) points to room for improvement, but this algorithm is certainly a better choice than Algorithms A_1 and C.

Table 3. Full-fit performance and cross validation (CV) statistics of Algorithm A_2. The coefficient of determination (R2) is based on log-transformed data. The Median Absolute Percentage Difference (MdPAD) is provided as percentage values and the Root Mean Square Difference (RMSD) in absolute units (mg m^{-3}). N is the number of observations.

		Full-fit			Cross validation		
Pigment	**N**	**Log[R^2]**	**RMSD (mg m^{-3})**	**MdAPD (%)**	**Log[CV.R^2]**	**CV.MdAPD (%)**	**CV.RMSD (mg m^{-3})**
butfuco	3320	0.501	0.035	44.914	0.500	45.212	0.034
hexfuco	3516	0.518	0.110	41.095	0.518	41.067	0.110
allo	1639	0.477	0.112	64.443	0.472	64.949	0.107
total_chl_a	3603	0.769	1.543	32.128	0.765	32.235	1.552
dv_chl_a	1020	0.342	0.083	37.971	0.336	38.089	0.061
total_chl_b	2970	0.585	0.082	51.849	0.582	51.775	0.079
fuco	3420	0.787	0.755	52.945	0.785	52.835	0.734
perid	2512	0.650	0.081	47.543	0.648	47.621	0.078
zea	3354	0.369	0.053	43.452	0.368	43.479	0.052

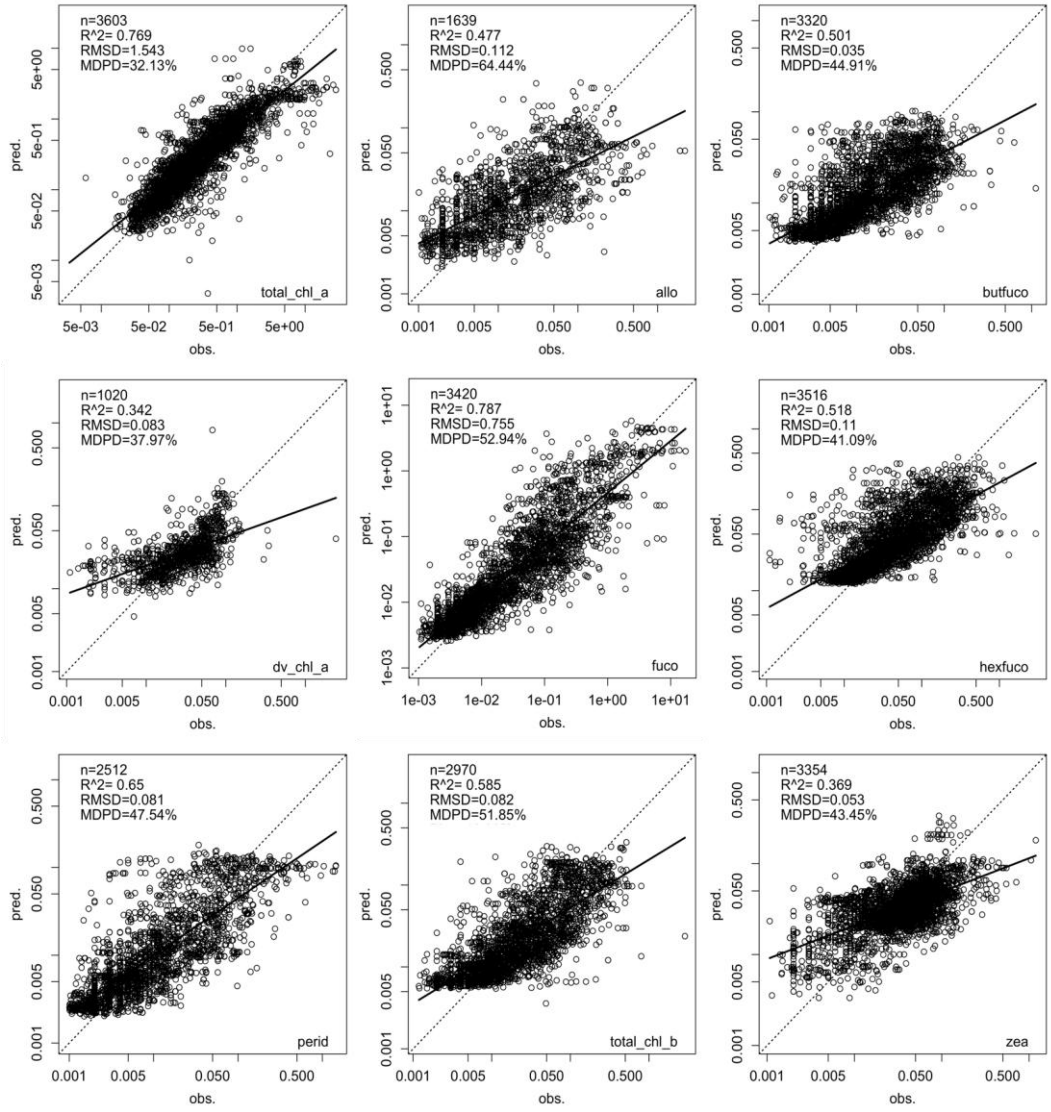

Figure 14. Algorithm A_2 (i.e., Pig-EOF-SST) full-fit model performance statistics and scatterplots of pigment concentration predictions and observation from match-up HPLC pigment data.

3.3 Algorithm B_1

3.3.1 Description

Algorithm B_1 is based on the deep-learning model by Li *et al.* (2023) (DL-PPCE model), which can estimate concentrations of 17 different phytoplankton pigments globally. The original DL-PPCE model is based on a global match-up dataset of R_{RS} at 412, 443, 469, 488, 531, 547, 555, 645, 667 and 678 from MODIS-Aqua (i.e., level-3 standard products at 4-km resolution) and *in situ* phytoplankton pigment observations measured by HPLC. The model adopts a fusion architecture of residual and pyramid networks to achieve robust estimation performance. The model inputs include three different data types: essential ocean colour variables, satellite-derived environmental variables and the slope of R_{RS}. Li *et al.* (2023) compared their model performances with various input variables to determine the most effective inputs. The results showed that R_{RS} in the essential ocean colour variables and SST in the environmental variables were the most critical input data. For Algorithm B_1, the model is established based on the *in situ* – match-up phytoplankton carbon dataset selected for algorithm development (see for details PVASR) and the output focused on the same nine

phytoplankton pigments as listed for Algorithms A_1 and A_2 (Table 1). Note that the DL-PPCE model of PHYTO-CCI is thus trained on six bands of OC-CCI matched with nine pigments from *in situ* observations.

3.2.2 Algorithm definition

Input data

The input data required for Algorithm B_1 are the same as for Algorithm A_1, R_{RS} at 412, 443, 490, 560, 620 and 665 nm from OC-CCI version 6, matched to *in situ* observations of phytoplankton pigments measured by HPLC. For the DL-PPCE model of PHYTO-CCI, nine pigments were considered for robust predictions (see Table 1). The developed algorithms for each pigment can be applied to $R_{RS}(\lambda)$ of OC-CCI to produce global maps of the nine identified pigments.

Retrieval algorithm

Algorithm B_1 is based on the DL-PPCE method (Li *et al.*, 2023), which simultaneously analyses different datatypes and obtains optimal outputs by continuously optimising the weight coefficients with fast calculation speed (Li *et al.*, 2020). Figure 15 provides a schematic overview of the DL-PPCE model as implemented within Algorithm B_1.

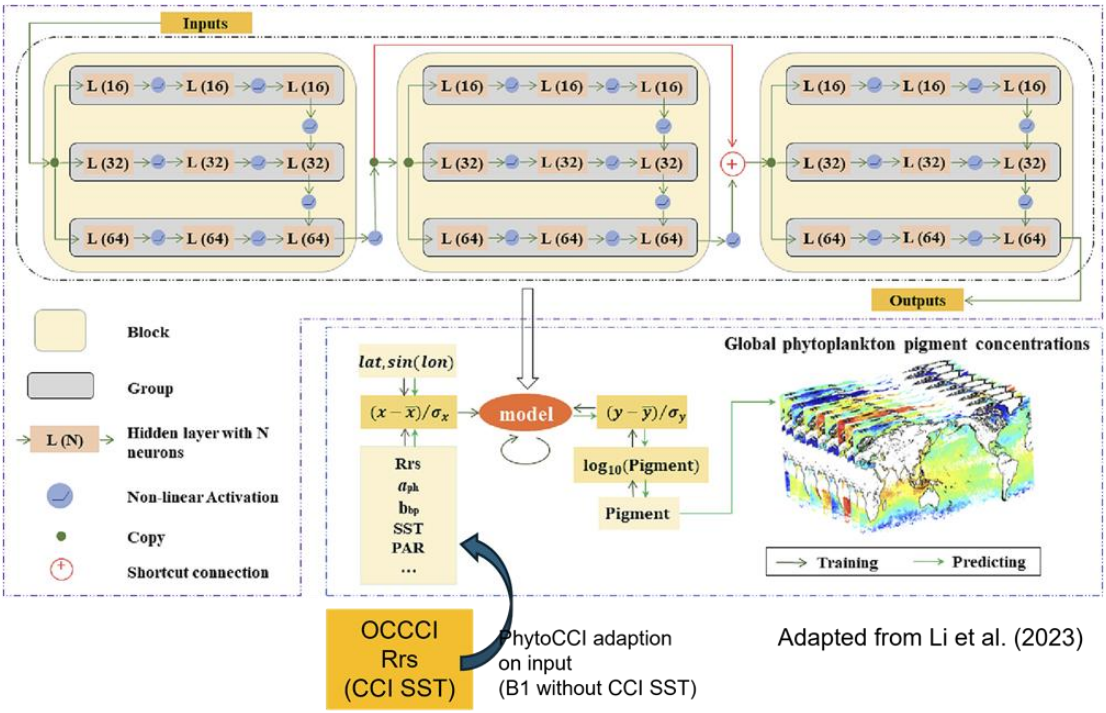

Figure 15. Schematic overview of the DL-PPCE as implemented for Algorithms B_1 and B_2 (with SST as additional input).

The algorithm builds on residual learning from the input data to stacked layers in the neural network (NN), introduced by He *et al.* (2016) as residual network (ResNet). This method redefines the layers as residual functions are learning about their inputs and applies this residual learning to every few stacked layers that form residual blocks in the NN. For estimating phytoplankton pigment concentrations from $R_{RS}(\lambda)$, complicated nonlinear function approximations are achieved as much as possible by increasing the depth of the NN and therefore using residual learning to guarantee the estimation accuracy. In addition, each pigment model adopts a pyramid network structure in each block, which employs groups with three different depths. This topological structure of the pyramid network structure, which is commonly used for redundancy reduction, facilitates multi-dimensional

feature learning on the different match-ups of phytoplankton pigment concentrations without sacrificing representational power and speed (as demonstrated in Yang *et al.*, 2018).

As mentioned above, DL-PPCE models combine two different learning frameworks: the ResNet structure and the pyramid network framework. As shown in Figure 15, the entire network consists of three network blocks in series, with residual connections added between the first and third blocks. To form each network block, three groups are created, all of which have three hidden layers to form a pyramid network, and the number of neurons in the hidden layers in different groups is 16, 32 and 64. Each group's outputs concatenate to the next group's third layer. In other words, the inputs of the third layer of each group include the outputs from the previous layer of the same group and the outputs of the previous group. Therefore, the three groups of fully connected networks have a specific parallel relationship and the outputs of the last group are used as the outputs of each block. Unlike the original ResNet models, for the implementation of the algorithm by Li *et al.* (2023), the blocks are redefined as residual functions learning about their inputs instead of redefining the layers. Using this combined neural network architecture, problems of structure redundancy, expensive runtime, overfitting and gradient diffusion can be avoided. Moreover, one advantage of this deep-learning architecture is handling data of varying lengths, allowing the DL-PPCE models to estimate multiple pigment concentrations with different numbers of match-up datapoints.

The detailed phytoplankton pigment retrievals procedure is as follows. The models use log-10-transformed estimated pigment concentrations as the target values to approach the centralised distribution of pigment concentrations because the matched *in situ* pigments followed a power log-normal distribution (see PSAVR). Additionally, the sine function of longitude, instead of longitude, is taken to ensure spatial continuity of inputs because 180°E and 180°W are the same lines. For input, data normalisation is necessary to improve the numerical stability of the model, each input is therefore standardised by subtracting the mean and then dividing by its standard deviation.

Output data

Similar to the other algorithms, Algorithm B_1 is able to predict nine phytoplankton pigment concentrations (in mg m^{-3}) at 4 km spatial resolution for each month between 1997 and 2025. An example of the monthly products for fucoxanthin, 19'-hexanoylfucoxantin, total chlorophyll and zeaxanthin is shown in Figure 16.

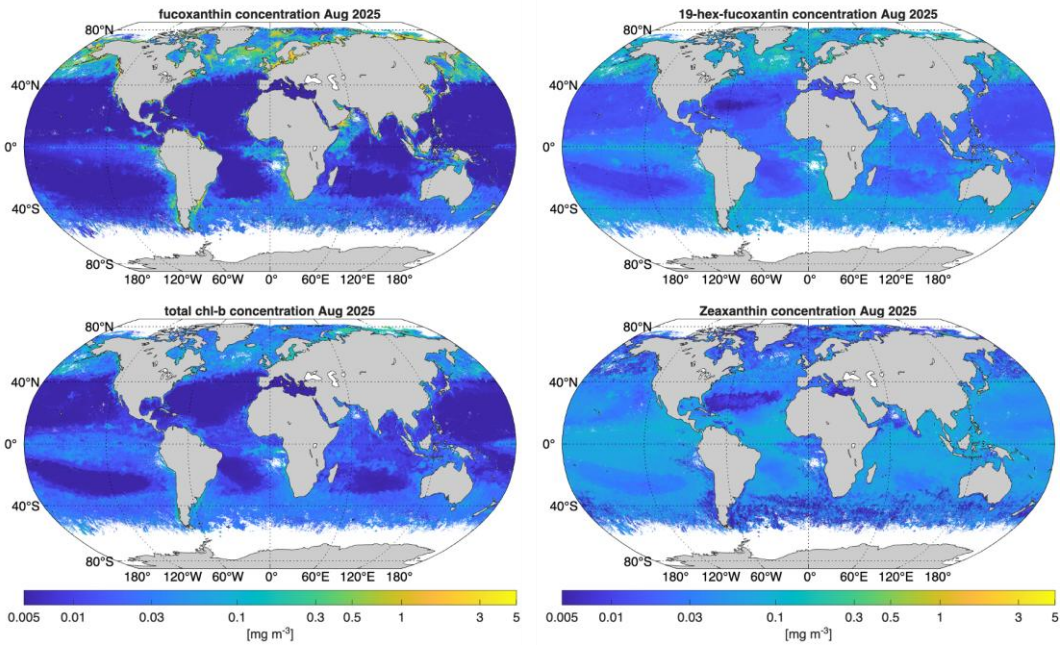

Figure 16. Example (August 2025) of monthly phytoplankton pigment product from Algorithm B_1 (Li et al.*, 2023; R_{RS} only) for the pigments fucoxanthin, 19'-hexanoylfucoxanthin, total chlorophyll-b (total chl-b) and zeaxanthin.*

3.3.3 Quality assessment

The estimation of phytoplankton pigment concentrations using Algorithm B_1 was validated against *in situ* observations measured by HPLC using a 10-fold Cross Validation (CV) method. This method was also used for evaluation of the original algorithm (Li *et al.*, 2023). Overall, results were robust for all nine phytoplankton pigments targeted in PHYTO-CCI. Using the same statistical metrics as in the CV of Algorithms A_1 and A_2, the Log[CV.R^2] and CV.RMSD were calculated (also available for the original algorithm of Li *et al.*, 2023). In addition, the mean absolute bias over all single CV statistics (CV.Bias) was calculated as:

$$Bias = \frac{\sum_{i=1}^{N}(c_{pi}-c_{oi})}{N}. \quad (3.3.1)$$

Scatter plots of the cross validation including statistics for Algorithm B_1 are presented in Figure 17. In addition, results for the CV statistics for Algorithm B_1 (and B_2) and the original algorithm of Li *et al.* (2023) are presented in Table 4, which also highlights where improved performance for Algorithm B_1 is achieved as compared to the original algorithm.

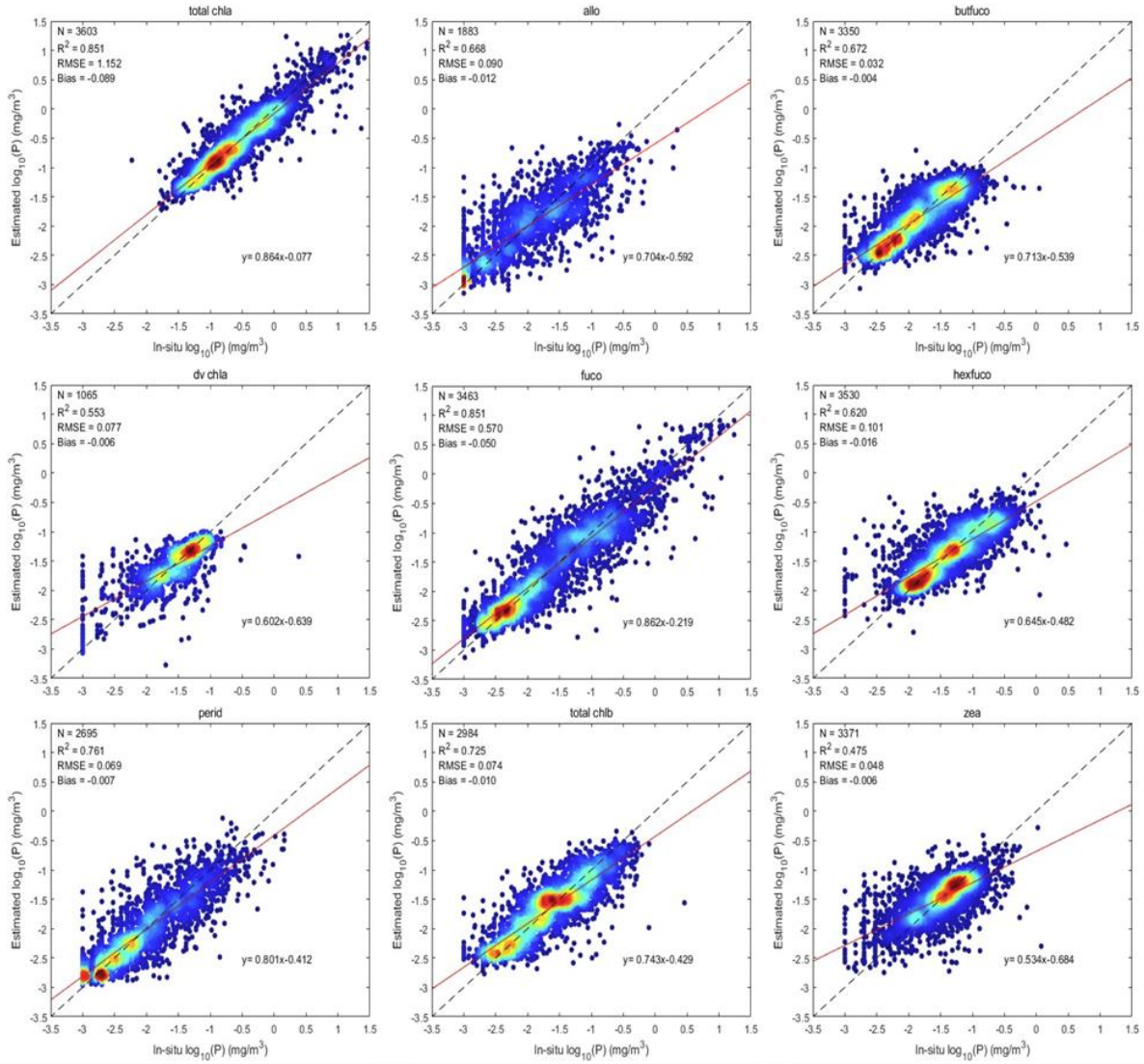

Figure 17. Cross validation statistics and scatterplots for the pigment predictions of Algorithm B_1 compared with the in situ *observation measured by HPLC for the PHYTO-CCI algorithm development dataset.*

Algorithm B_1 performed well with high values of Log[CV.R^2] (0.62 to 0.86) for six out of eight pigments (i.e., excluding chlorophyll-a), with marginaly lower performance in some pigments compared with the original algorithm (i.e., for 19'-hexanoylfucoxanthin, alloxanthin and fucoxanthin) and better performance for 19'-butanoylfucoxanthin, total chlorophyll-b and peridinin (Table 3). However, the Log[CV.R^2] were lower for divinyl chlorophyll-a and zeaxanthin compared with the original model, decreasing by 0.23 and 0.16, respectively. Considering CV.RMSD, the values were worse for fucoxanthin, divinyl chlorophyll-a and alloxanthin, increasing from 0.46 to 0.57 mg m^{-3}, 0.02 to 0.08 mg m^{-3}, and 0.04 to 0.09 mg m^{-3}, respectively, but improved for 19'-butanoylfucoxanthin and peridinin, decreasing from 0.19 to 0.03 mg m^{-3} and 0.15 to 0.07 mg m^{-3}, respectively, and similar for 19'-hexanoylfucoxanthin total chlorophyll-b and zeaxanthin. Overall, the performance of Algorithm B_1 can be considered to be high, confirming that even with using less spectral bands (6 bands from OC-CCI versus 10 from MODIS-Aqua), similarly robust results were obtained than that of the original algorithm by Li *et al.* (2023). In addition, compared with Algorithms A_1 and A_2, the CV statistics showed higher correlation (~+0.1 for Log[CV.R^2]) and nearly an order of magnitude lower CV.RMSD.

Table 4. Cross validation statistics for the original DL-PPCE algorithm by Li et al. *(2023) and the implementation in PHYTO-CCI: Algorithm B_1 that is based on R_{RS} only and Algorithm B_2 that additionally includes SST. The coefficient of determination (R^2), bias and Root Mean Square Difference (RMSD) are provided. Red (blue) highlights where the original algorithm showed poorer performance compared with Algorithms B_1 and B_2 (B_2 only) for RMSD and Log[R^2]. Magenta highlights where Algorithm B^1 performed better than Algorithm B_2 for a specific statistical metric. The number of match-up datapoints (N) compared is the same for Algorithms B_1 and B_2.*

	Original algorithm			Algorithm B_1			Algorithm B_2		
Pigment	**Log[R^2]**	**RMSD (mg m^{-3})**	**N**	**Log[R^2]**	**RMSD (mg m^{-3})**	**Bias (mg m^{-3})**	**Log[R^2]**	**RMSD (mg m^{-3})**	**Bias (mg m^{-3})**
butfuco	0.42	0.19	3350	0.67	0.03	-0.004	0.71	0.03	-0.004
hexfuco	0.66	0.09	3530	0.62	0.10	-0.016	0.68	0.10	-0.012
allo	0.78	0.04	1883	0.67	0.09	-0.012	0.71	0.09	-0.012
total_chl_a			3603	0.85	1.15	-0.089	0.86	1.10	-0.092
dv_chl_a	0.78	0.02	1065	0.55	0.08	-0.006	0.67	0.08	-0.005
total_chl_b	0.63	0.07	2984	0.73	0.07	-0.010	0.75	0.07	-0.009
fuco	0.87	0.46	3463	0.86	0.57	-0.050	0.88	0.60	-0.050
perid	0.71	0.15	2695	0.76	0.07	-0.007	0.78	0.07	-0.006
zea	0.64	0.05	3371	0.48	0.05	-0.006	0.65	0.05	-0.006

3.4 Algorithm B_2

3.4.1 Description

As for Algorithm B_1, Algorithm B_2 is based on the deep-learning model of Li *et al.* (2023) to predict 17 different phytoplankton pigments at the global scale. The original DL-PPCE model was based on R_{RS} values at 10 different wavebands from MODIS-Aqua that were matched in space and time to *in situ* observations of pigments measured by HPLC. The model adopts a fusion architecture of residual and pyramid networks to achieve robust estimation performance. The model inputs include essential ocean colour variables, satellite-derived environmental variables and the slope of R_{RS}. Through intercomparison of various input data configurations, Li *et al.* (2023) established that the model performed best when R_{RS} and SST were included as input variables. For PHYTO-CCI, the model is retrained as Algorithm B_2 using the *in situ* – satellite match-up dataset for algorithm development as described in the PVASR. In contrast to Algorithm B_1, Algorithm B_2 incorporates the additional variable SST for improved prediction of the nine pigments targeted in PHYTO-CCI.

3.4.2 Algorithm definition

Input data

The input data required for Algorithm B_1 is similar to that needed for Algorithm A_2, and includes R_{RS} at 412, 443, 490, 560, 620 and 665 nm from OC-CCI version 6 and SST from SST-CCI version 3 (i.e., available through CRS), matched to *in situ* observations of phytoplankton pigments measured by HPLC. The algorithm is trained to predict the nine phytoplankton pigments listed in Table 1, i.e., the same as for Algorithms A_1, A_2 and B_1. After the development of the algorithms for each phytoplankton pigment, the algorithms can be applied to gridded data of OC-CCI version 6 and SST-CCI version 3, provided that they are at the same spatial and temporal resolution.

Retrieval algorithm

Algorithm B_2 is based on the same method as described for Algorithm B_1 and we refer to Section 3.2.2 for the details of this algorithm. Briefly, the algorithm is based on a neural network approach and combines a ResNet structure (He et al., 2016) with the pyramid network framework (Yang *et al.*, 2018). For the implementation of the algorithm by Li *et al.* (2023), the blocks in the ResNet structure are redefined as residual functions learning about their inputs instead of redefining the layers. Problems of structure redundancy, expensive runtime, overfitting and gradient diffusion can be avoided using this combined neural network architecture. It also allows handling of data of varying lengths, allowing the DL-PPCE models to estimate multiple pigment concentrations with different numbers of match-up data points. Figure 15 provides a schematic overview of the DL-PPCE model by Li *et al.* (2023) as implemented within Algorithm B_2 (i.e., with the additional input of SST). Note that for algorithm B_2, SST has been standardised in the same way as the other input variables (as described for Algorithm A_1), i.e., SST is standardised by subtracting the mean and the mean is then divided by the standard deviation.

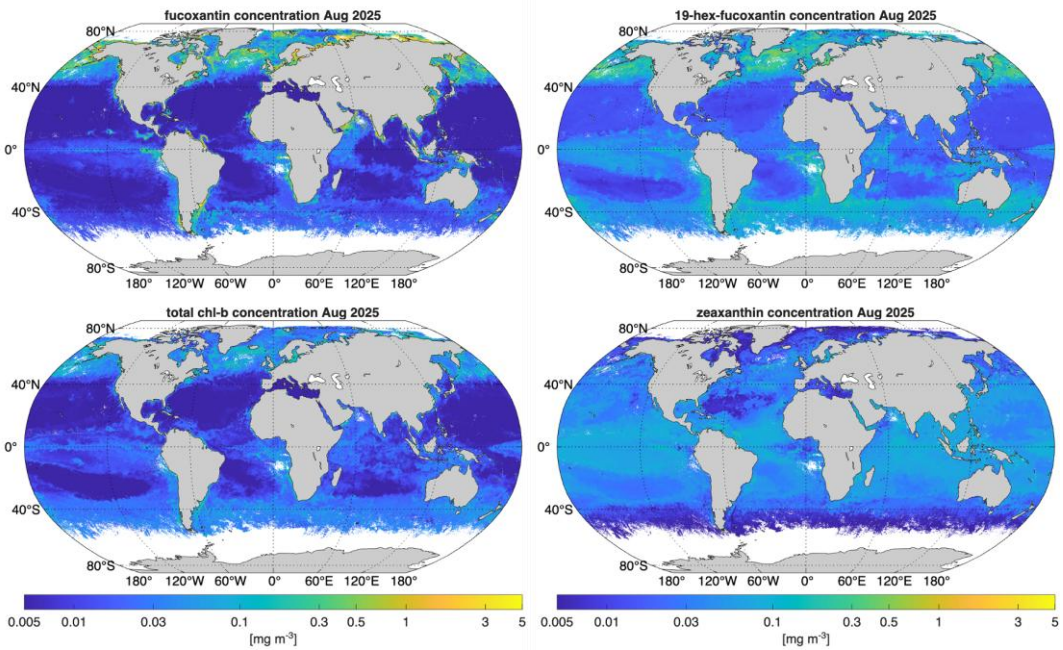

Figure 18. Example (August 2025) of monthly phytoplankton pigment product from Algorithm B_2 (Li et al.*, 2023; R_{RS} + SST) for the pigments fucoxanthin, 19'-hexanoylfucoxanthin, total chlorophyll-b (total chl-b) and zeaxanthin.*

Output data

Algorithm B2 estimates eight phytoplankton pigments and the total chlorophyll-a concentration (in mg m^{-3}) at 4 km spatial resolution for each month between 1997 and 2025. An example of the

monthly products for fucoxanthin, 19'-hexanoylfucoxantin, total chlorophyll and zeaxanthin is shown in Figure 18.

3.4.3 Quality assessment

As for the evaluation of the original (Li *et al.*, 2023) and B_1 algorithms, the estimation of phytoplankton pigment concentrations by Algorithm B_2 was compared with *in situ* pigment observations measured by HPLC using a 10-fold Cross Validation (CV) method. Overall, results were robust for all eight phytoplankton pigments and the total chlorophyll-a concentration based on the same statistical metrics used for the CV of the original algorithm and Algorithm B_1 (i.e., Log10[CV.R^2], CV.RMSD and CV.Bias). Scatter plots of the comparison between estimated and observed pigments, including CV statistics, for Algorithm B_2 are presented in Figure 19. In addition, the CV statistics for Algorithm B_2 along with those of the original algorithm and Algorithm B_1 are presented in Table 4, which also highlights where improved performance for Algorithms B_1 and B_2, and for Algorithm B_2 specifically, are achieved as compared with the original algorithm.

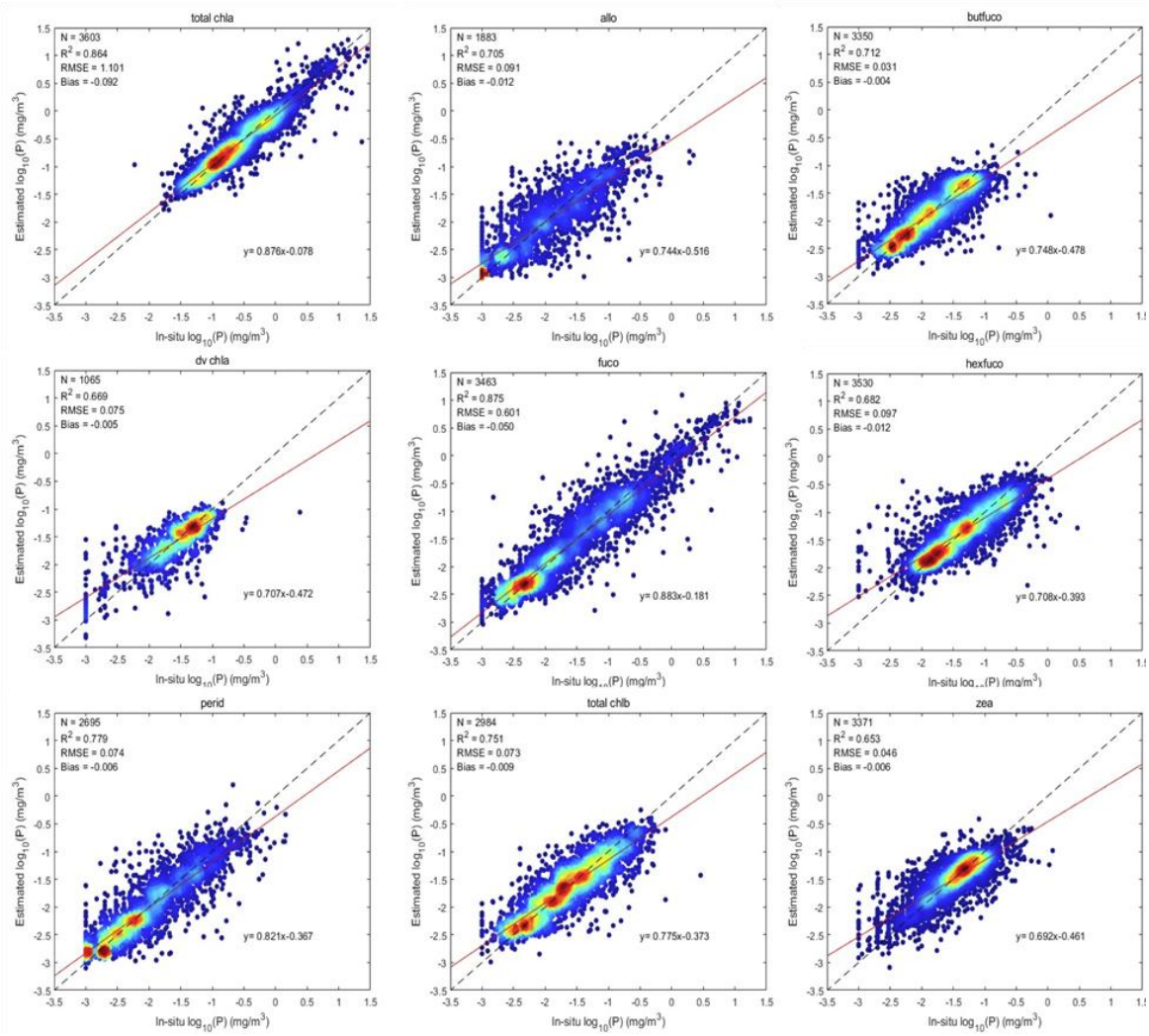

Figure 19. Cross validation statistics and scatterplots for the pigment predictions of Algorithm B_2 compared with the in situ *observation measured by HPLC for the PHYTO-CCI algorithm development dataset.*

Results showed that the performance of Algorithm B_2 was better than Algorithm B_1 when Log[CV.R^2] was considered and performance was similar based on CV.Bias and CV.RMSD, except for fucoxanthin (Table 4). The CV.RMSD for fucoxanthin was worse in Algorithm B_2 (0.60 mg m^{-3}) compared with Algorithm B_1 (0.57 mg m^{-3}) and considerably higher than in the original algorithm by

Li *et al.* (2023) (0.46 mg m^{-3}). Algorithm B_2 showed higher Log[CV.R^2] (i.e., between 0.65 and 0.88) than the original algorithm for all phytoplankton pigments, except for divinyl chlorophyll-a (0.67 versus 0.78) and alloxanthin (0.71 versus 0.78). Considering CV.RMSD, similar to Algorithm B_1, the values are worse for fucoxanthin, divinyl chlorophyll-a and alloxanthin compared with the original algorithm (i.e., increasing from 0.46 to 0.60 mg m^{-3}, 0.02 to 0.08 mg m^{-3} and 0.04 to 0.09 mg m^{-3}, respectively), but improved for 19'-butanoylfucoxanthin and peridinin (i.e., decreasing from 0.19 to 0.03 mg m^{-3} and 0.15 to 0.07 mg m^{-3}, respectively) and similar for zeaxanthin, 19'-hexanoylfucoxanthin and total chlorophyll-b. Overall, the performance of Algorithm B_2 can be treated as very high and confirms that adding SST as an input variable, even with less spectral information, provides similar robust results and generally better results compared with the original algorithm by Li *et al.* (2023). As for Algorithm B_1, compared with Algorithms A_1 and A_2, the CV statistics of Algorithm B_2 showed higher correlation (~+0.1 for Log[CV.R^2]) and nearly an order of magnitude lower CV.RMSD. These good CV statistics were also reflected in the overall algorithm performance within the independent validation as documented in the PVASR.

3.5 Algorithm C

3.5.1 Description

Algorithm C is also a machine learning-based model, developed following opportunities highlighted by earlier algorithm development and recommendations on retrieval of pigments from space. There are several satellite-retrieval algorithms available for the estimation of phytoplankton diversity metrics, including phytoplankton pigments and phytoplankton functional types. As described in Sections 3.1 and 3.2, Bracher *et al.* (2015) used empirical orthogonal functions to predict pigment concentrations from satellite-derived remote-sensing reflectance and *in situ* hyperspectral reflectance. This method was further developed by Xi *et al.* (2020, 2021) for application at the global scale using multispectral satellite observations, as described for Algorithms A_1 and A_2. El-Hourany *et al.* (2019) reported high accuracy in the retrieval of all secondary pigments from satellite-derived remote sensing reflectances using Self-Organising Maps, trained on *in situ* pigment observations from HPLC measurements and optical information at four SeaWiFS wavebands as well as the total chlorophyll-a concentration and SST. Along those recent developments, Stock & Subramiam (2020) conducted a comparison of machine-learning models using a spatial-block cross-validation strategy. They reported significant accuracy for retrieval of chlorophyll-a, fucoxanthin and zeaxanthin and proposed a random forest approach as one of the best performing machine learning models to retrieve pigments from space. Following these earlier approaches, the algorithm by Laurenson *et al.* (in preparation) was developed for the ESA BiCOME and BOOMS projects using a random forest model.

3.5.2 Algorithm definition

Input data

The input data required for the training of Algorithm C is similar to that of the other phytoplankton algorithms described, i.e., the R_{RS} from OC-CCI matched to *in situ* observations of phytoplankton pigments from HPLC measurements. However, Algorithm C was not retrained for the first development round of PHYTO-CCI and an existing database of *in situ* pigment observations was used (33,640 data points; Sun *et al.*, 2023). This *in situ* data was matched in space and time to observations of OC-CCI version 6 at daily, 4-km resolution. Quality control was applied to the *in situ* – satellite match-up dataset following standard procedures, i.e., satellite data was extracted for a 3x3 pixel window and if more than 50% of pixels were invalid and/or if the coefficient of variation was higher than 0.15 (i.e., data with anomalous variability around the mean), data were excluded from further analysis. Only *in situ* observations from depths less than 5 meter were used. The match-up dataset consisted of 3,291 datapoints, of which approximately a third was dedicated to validation of

the trained algorithm (i.e., data from 2002, 2007, 2012 and 2017). We note that while this *in situ* – satellite match-up dataset is not exactly the same as the one used in PHYTO-CCI, much of the training dataset used for Algorithm C (Sun *et al.,* 2023) is contained in the algorithm development and intercomparison dataset of PHYTO-CCI.

Model development resulted in the selection of the following satellite data that are needed for prediction of eight phytoplankton carbon pigments (Table 1; excluding divinyl chlorophyll-a):

- Remote sensing reflectance (R_{RS}) at 412, 443, 490, 510, 560 and 665 nm from OC-CCI version 6
- Bathymetry from ETOPO 1 (NOAA, 2009; Amante & Eakins, 2009)

For global application of Algorithm C in PHYTO-CCI, the input data are required a 4 km spatial and monthly temporal resolution.

Retrieval algorithm

Algorithm C is based on a Random Forest (RF) model (Breiman, 2001), which is an ensemble learning method that creates multiple decision trees that can capture non-linear relationships and is relatively robust to overfitting and noise, making it suitable for small and noisy datasets (Opitz & Maclin, 1999). The RF model was trained to predict the concentrations of eight pigments (Table 1; excluding divinyl chlorophyll-a) and all variables were log-transformed to better represent both low and high concentrations of specific pigments. Model hyperparameters, including the number of trees, number of features, number of training samples and maximum tree depth were optimised using a grid search, and optimised separately for each pigment.

A peer-reviewed publication of Algorithm C is currently being prepared for submission and some further details are available in the technical reports of the BOOMs project (Martinez-Vicente *et al.*, 2025).

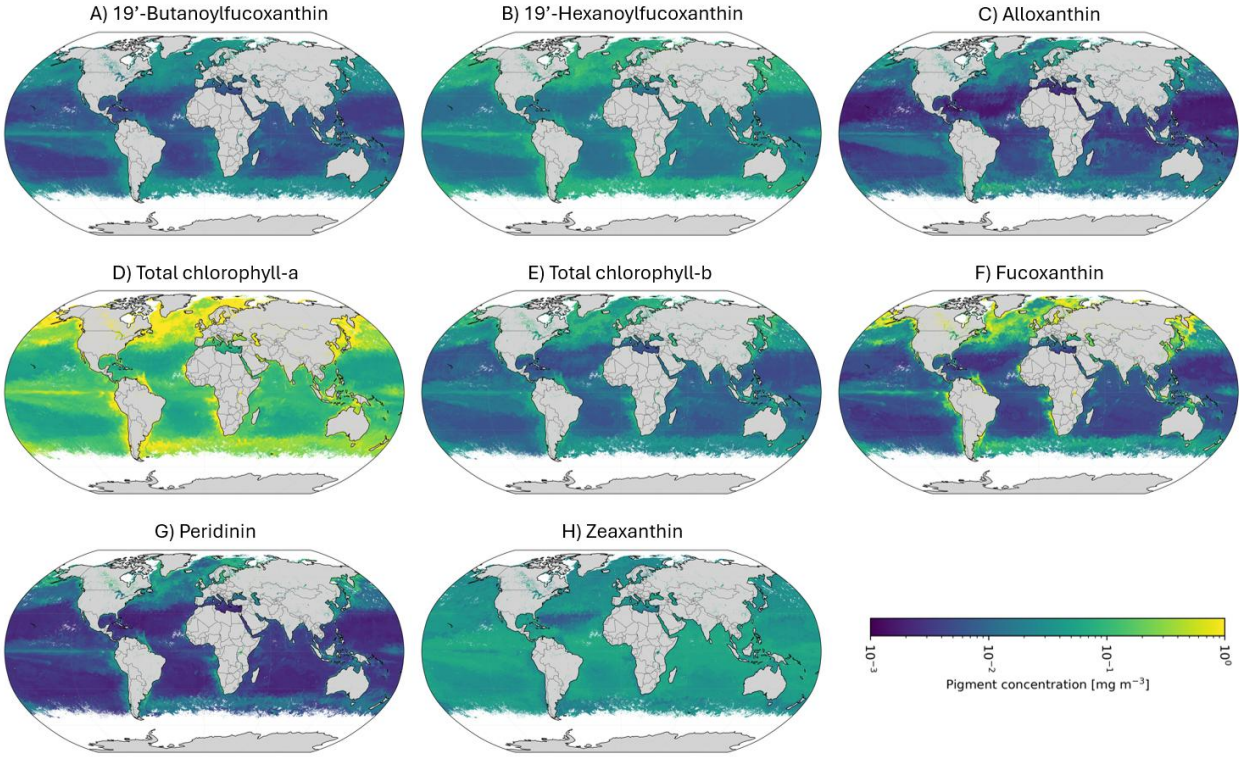

Figure 20. Example (May 2010) of the eight phytoplankton pigments estimated using Algorithm C.

Output data

Algorithm C estimates eight phytoplankton pigments, including 19'-buthexanoylfucaoxanthin, 19'-hexanoylfucoxanthin, alloxanthin, total chlorophyll-a, total chlorophyll-b, fucoxanthin, peridinin and zeaxanthin at 4 km spatial resolution and monthly temporal resolution between 1997 and 2025 (Figure 20). Compared with other algorithms in PHYTO-CCI, Algorithm C does not estimate divinyl chlorophyll-a.

3.5.3 Quality assessment

Algorithm C showed encouraging results based on the training dataset with good predictive performance for chlorophyll-a, fucoxanthin and peridinin, but performance was lower for some other pigment such as zeaxanthin. The performance of the algorithm was further assess using *in situ* observations of phytoplankton pigments that were not used for training, specifically of the years 2002, 2007, 2012 and 2017. Preliminary results of this analysis showed that the coefficient of determination was higher for the total chlorophyll-a concentration, which outperformed estimates from OC-CCI for this pigment, and fucoxanthin (Figure 21). Intermediate performance of Algorithm C was observed for total chlorophyll-b and 19'-hexanoylfucoxanthin and lower performance was observed for 19'-butanoylfucoxanthin, peridinin, alloxantin and zeaxanthin (Figure 21). Further analysis will focus on the estimation of bias and variance. Overall, the performance of this algorithm in comparison with the other phytoplankton pigment algorithms shows that there is scope for further development and improvement of Algorithm C, noting that this algorithm was trained on a different subset of the *in situ* – satellite match-up database.

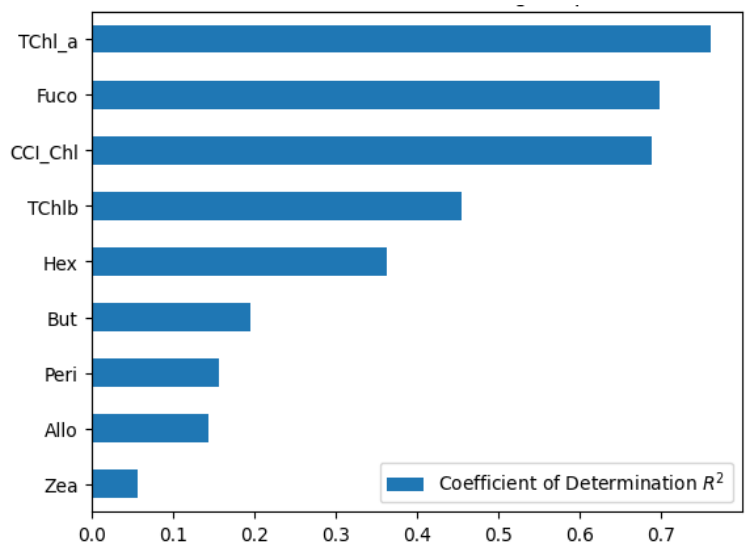

Figure 21. Coefficient of determination (R^2) of Algorithm C in comparison with in situ *pigment observations that were not used for training of the algorithm. Note that OC-CCI version 6.0 chlorophyll estimate is presented as CCI_Chl for comparison. The statistic is computed on log-10 transformed data.*

4 Algorithm blending

An innovation that was introduced by OC-CCI, and subsequently adopted by others such as Lakes CCI, was to identify best-performing algorithms for each optical water class, and then blend algorithms according to the membership of each class in each pixel (Jackson *et al.* 2017a). Algorithm blending is desirable for ECV products based on ocean-colour observations as typically a number of satellite-based algorithms exist for the estimation of the targeted ECV product and each algorithm has geographical areas and water column conditions in which it performs well or poorly. It has been proposed by Moore *et al.* (2001, 2014) that fuzzy optical class memberships might be suitable for

the optimal blending of such algorithms. A successful algorithm blending approach should combine algorithms seamlessly to produce the optimal resultant product without introducing boundary artefacts. Jackson *et al.* (2017a) demonstrated the application of fuzzy class membership to product blending for chlorophyll-a for application in OC-CCI. Using the OCI, OC3 and OC5 algorithms (Gohin *et al.* 2002; Hu *et al.* 2012; O'Reilly *et al.* 2000), the optimal chlorophyll-a algorithm per optical water class was identified by a round robin assessment utilising a quantitative scoring system (Brewin *et al.* 2015). The optimal chlorophyll-a algorithm differed across water classes with OCI performing best for water classes 1-7 (clearer ocean waters), OC3 for water classes 13-14 (more turbid, coastal waters) and OC5 for water classes 8-12 (generally transitional regions between open ocean and near coastal regions). The optimal algorithms were then blended using a weighting scheme based on the membership of each optical water class for each pixel. Results in the eastern tropical Atlantic Ocean, chosen as an example region which contains a range of dominant optical water classes, showed that the blended chlorophyll-a product in the oligotrophic gyre is essentially identical to the OCI algorithm (see Figure 8 from Jackson *et al.* 2017a). Conversely, the blended chlorophyll-a estimate in the coastal waters is closer to the OC3 and OC5 algorithms. Importantly, the method does not introduce artificial boundaries in the final blended product. In PHYTO-CCI, we plan to use a similar algorithm blending approach for phytoplankton carbon and phytoplankton pigments, depending on the outcomes of the algorithm intercomparison. It is important to note that when optical water classes – or other classification approaches – are used for algorithm blending, there is a requirement that there is sufficient *in situ* data in each class, to bring confidence to the algorithm selection (Jackson *et al.* 2017a).

5 References

Aden, A. L., & Kerker, M. (1951). Scattering of Electromagnetic Waves from Two Concentric Spheres. Journal of Applied Physics, 22(10), 1242–1246. https://doi.org/10.1063/1.1699834

Alvain, S., Moulin, C., Dandonneau, Y., & Bréon, F. (2005). Remote sensing of phytoplankton groups in case 1 waters from global SeaWiFS imagery. Deep Sea Research Part I Oceanographic Research Papers, 52(11), 1989–2004. https://doi.org/10.1016/j.dsr.2005.06.015

Alvain, S., Moulin, C., Dandonneau, Y., & Loisel, H. (2008). Seasonal distribution and succession of dominant phytoplankton groups in the global ocean: A satellite view. Global Biogeochemical Cycles, 22(3). https://doi.org/10.1029/2007gb003154

Amante, C., Eakins, B.W. (2009). ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC-24. National Geophysical Data Center, NOAA. https://doi.org/10.7289/v5c8276m

Bader, H. (1970). The hyperbolic distribution of particle sizes. Journal of Geophysical Research Atmospheres, 75(15), 2822–2830. https://doi.org/10.1029/jc075i015p02822

Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochemical Cycles, 19(1). https://doi.org/10.1029/2004gb002299

Bellacicco, M., Volpe, G., Colella, S., Pitarch, J., & Santoleri, R. (2016). Influence of photoacclimation on the phytoplankton seasonal cycle in the Mediterranean Sea as seen by satellite. Remote Sensing of Environment, 184, 595–604. https://doi.org/10.1016/j.rse.2016.08.004

Bellacicco, M., Volpe, G., Briggs, N., Brando, V., Pitarch, J., Landolfi, A., Colella, S., Marullo, S., & Santoleri, R. (2018). Global distribution of non-algal particles from ocean color data and implications for phytoplankton biomass detection. Geophysical Research Letters, 45(15), 7672–7682. https://doi.org/10.1029/2018gl078185

Bellacicco, M., Cornec, M., Organelli, E., Brewin, R. J. W., Neukermans, G., Volpe, G., Barbieux, M., Poteau, A., Schmechtig, C., D'Ortenzio, F., Marullo, S., Claustre, H., & Pitarch, J. (2019). Global Variability of Optical Backscattering by Non-algal particles From a Biogeochemical-Argo Data Set. Geophysical Research Letters, 46(16), 9767–9776. https://doi.org/10.1029/2019gl084078

Bellacicco, M., Pitarch, J., Organelli, E., Martinez-Vicente, V., Volpe, G., & Marullo, S. (2020). Improving the Retrieval of Carbon-Based Phytoplankton Biomass from Satellite Ocean Colour Observations. Remote Sensing, 12(21), 3640. https://doi.org/10.3390/rs12213640

Bernard, S., Probyn, T. A., Quirantes, A. (2009). Simulating the optical properties of phytoplankton cells using a two-layered spherical geometry. Biogeosciences Discussions, 6, 1497–1563. https://doi.org/10.5194/bgd-6-1497-2009.

Boss, E., Twardowski, M. S., & Herring, S. (2001). Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. Applied Optics, 40(27), 4885. https://doi.org/10.1364/ao.40.004885

Bracher, A., Taylor, M. H., Taylor, B., Dinter, T., Röttgers, R., & Steinmetz, F. (2015). Using empirical orthogonal functions derived from remote-sensing reflectance for the prediction of phytoplankton pigment concentrations. Ocean Science, 11(1), 139–158. https://doi.org/10.5194/os-11-139-2015

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brewin, R. J. W., Devred, E., Sathyendranath, S., Lavender, S. J., & Hardman-Mountford, N. J. (2011). Model of phytoplankton absorption based on three size classes. Applied Optics, 50(22), 4535. https://doi.org/10.1364/ao.50.004535

Brewin, R. J., Sathyendranath, S., Jackson, T., Barlow, R., Brotas, V., Airs, R., & Lamont, T. (2015). Influence of light in the mixed-layer on the parameters of a three-component model of phytoplankton size class. Remote Sensing of Environment, 168, 437–450. https://doi.org/10.1016/j.rse.2015.07.004

De Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. Journal of Geophysical Research Atmospheres, 109(C12). https://doi.org/10.1029/2004jc002378

Dennison, P. E., Halligan, K. Q., & Roberts, D. A. (2004). A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. Remote Sensing of Environment, 93(3), 359–367. https://doi.org/10.1016/j.rse.2004.07.013

El Hourany, R., Saab, M. A., Faour, G., Aumont, O., Crépon, M., & Thiria, S. (2019). Estimation of secondary phytoplankton pigments from satellite observations using Self-Organizing Maps (SOMs). Journal of Geophysical Research Oceans, 124(2), 1357–1378. https://doi.org/10.1029/2018jc014450

Geider, R. J. (1987). Light and temperature dependence of the carbon to chlorophyll a ratio in microalgae and cyanobacteria: Implications for physiology and growth of phytoplankton. New Phytologist, 106(1), 1–34. https://doi.org/10.1111/j.1469-8137.1987.tb04788.x

Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1996). A dynamic model of photoadaptation in phytoplankton. Limnology and Oceanography, 41(1), 1–15. https://doi.org/10.4319/lo.1996.41.1.0001

Geider, R., MacIntyre, H., & Kana, T. (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. https://doi.org/10.3354/meps148187

Geider, R. J., MacIntyre, H. L., & Kana, T. M. (1998). A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. Limnology and Oceanography, 43(4), 679–694. https://doi.org/10.4319/lo.1998.43.4.0679

Gohin, F., Druon, J. N., & Lampert, L. (2002). A five channel chlorophyll concentration algorithm applied to SeaWiFS data processed by SeaDAS in coastal waters. International Journal of Remote Sensing, 23(8), 1639–1661. https://doi.org/10.1080/01431160110071879

He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 770–778. https://doi.org/10.1109/ CVPR.2016.90.

Hendrycks, D., Lee, K., & Mazeika, M. (2019). Using Pre-Training can improve model robustness and uncertainty. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1901.09960

Hirata, T., Hardman-Mountford, N. J., Brewin, R. J. W., Aiken, J., Barlow, R., Suzuki, K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., & Yamanaka, Y. (2011). Synoptic relationships between surface Chlorophyll- a and diagnostic pigments specific to phytoplankton functional types. Biogeosciences, 8(2), 311–327. https://doi.org/10.5194/bg-8-311-2011

Hu, C., Lee, Z., & Franz, B. (2011). Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. Journal of Geophysical Research Atmospheres, 117(C1). https://doi.org/10.1029/2011jc007395

Jackson, T., Sathyendranath, S., & Mélin, F. (2017a). An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. Remote Sensing of Environment, 203, 152–161. https://doi.org/10.1016/j.rse.2017.03.036

Jackson, T., Sathyendranath, S., & Platt, T. (2017b). An exact solution for modeling photoacclimation of the Carbon-to-Chlorophyll ratio in phytoplankton. Frontiers in Marine Science, 4.

https://doi.org/10.3389/fmars.2017.00283
Jonasz, M. (1983). Particle-size distributions in the Baltic. Tellus B, 35(5), 346. https://doi.org/10.3402/tellusb.v35i5.14624
Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2009). Retrieval of the particle size distribution from satellite ocean color observations. Journal of Geophysical Research Atmospheres, 114(C9). https://doi.org/10.1029/2009jc005303
Kostadinov, T. S., Siegel, D. A., & Maritorena, S. (2010). Global variability of phytoplankton functional types from space: assessment via the particle size distribution. Biogeosciences, 7(10), 3239–3257. https://doi.org/10.5194/bg-7-3239-2010
Kostadinov, T. S., Milutinović, S., Marinov, I., & Cabré, A. (2016). Carbon-based phytoplankton size classes retrieved via ocean color estimates of the particle size distribution. Ocean Science, 12(2), 561–575. https://doi.org/10.5194/os-12-561-2016
Kostadinov, T. S., Lain, L. R., Kong, C. E., Zhang, X., Maritorena, S., Bernard, S., Loisel, H., Jorge, D. S. F., Kochetkova, E., Roy, S., Jonsson, B., Martinez-Vicente, V., & Sathyendranath, S. (2023). Ocean color algorithm for the retrieval of the particle size distribution and carbon-based phytoplankton size classes using a two-component coated-sphere backscattering model. Ocean Science, 19(3), 703–727. https://doi.org/10.5194/os-19-703-2023
Kulk, G., Platt, T., Dingle, J., Jackson, T., Jönsson, B., Bouman, H., Babin, M., Brewin, R., Doblin, M., Estrada, M., Figueiras, F., Furuya, K., González-Benítez, N., Gudfinnsson, H., Gudmundsson, K., Huang, B., Isada, T., Kovač, Ž., Lutz, V., . . . Sathyendranath, S. (2020). Primary Production, an Index of Climate Change in the Ocean: Satellite-Based Estimates over Two Decades. Remote Sensing, 12(5), 826. https://doi.org/10.3390/rs12050826
Kulk, G., Platt, T., Dingle, J., Jackson, T., Jönsson, B. F., Bouman, H. A., Babin, M., Brewin, R. J. W., Doblin, M., Estrada, M., Figueiras, F. G., Furuya, K., González-Benítez, N., Gudfinnsson, H. G., Gudmundsson, K., Huang, B., Isada, T., Kovač, Ž., Lutz, V. A., . . . Sathyendranath, S. (2021). Correction: Kulk et al. Primary Production, an Index of Climate Change in the Ocean: Satellite-Based Estimates over Two Decades. Remote Sens. 2020, 12, 826. Remote Sensing, 13(17), 3462. https://doi.org/10.3390/rs13173462
Laws, E. A., & Bannister, T. T. (1980). Nutrient- and light-limited growth of Thalassiosira fluviatilis in continuous culture, with implications for phytoplankton growth in the ocean1. Limnology and Oceanography, 25(3), 457–473. https://doi.org/10.4319/lo.1980.25.3.0457
Lee, Z., Carder, K. L., & Arnone, R. A. (2002). Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Applied Optics, 41(27), 5755. https://doi.org/10.1364/ao.41.005755
Li, Q. P., Franks, P. J. S., Landry, M. R., Goericke, R., & Taylor, A. G. (2010). Modeling phytoplankton growth rates and chlorophyll to carbon ratios in California coastal and pelagic ecosystems. Journal of Geophysical Research Atmospheres, 115(G4). https://doi.org/10.1029/2009jg001111
Li, X., Li, X., Yang, Y., Ishizaka, J., Li, X., & Li, X. (2023). Global estimation of phytoplankton pigment concentrations from satellite data using a deep-learning-based model. Remote Sensing of Environment, 294, 113628. https://doi.org/10.1016/j.rse.2023.113628
Losa, S. N., Soppa, M. A., Dinter, T., Wolanin, A., Brewin, R. J. W., Bricaud, A., Oelker, J., Peeken, I., Gentili, B., Rozanov, V., & Bracher, A. (2017). Synergistic exploitation of Hyper- and Multi-Spectral Precursor sentinel measurements to determine phytoplankton functional types (SYNSENPFT). Frontiers in Marine Science, 4. https://doi.org/10.3389/fmars.2017.00203
Loisel, H., & Stramski, D. (2000). Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. Applied Optics, 39(18), 3001. https://doi.org/10.1364/ao.39.003001
Longhurst, A.R. (2007). Ecological Geography of the Sea, 2nd ed.; Elsevier Academic Press: Cambridge, MA, USA, p. 542.
Maniaci, G., Brewin, R. J. W., & Sathyendranath, S. (2022). Concentration and distribution of phytoplankton nitrogen and carbon in the Northwest Atlantic and Indian Ocean: A simple model with applications in satellite remote sensing. Frontiers in Marine Science, 9. https://doi.org/10.3389/fmars.2022.1035399
Marañón, E., Cermeño, P., Huete-Ortega, M., López-Sandoval, D. C., Mouriño-Carballido, B., & Rodríguez-Ramos, T. (2014). Resource supply overrides temperature as a controlling factor of marine phytoplankton growth. PLoS ONE, 9(6), e99312. https://doi.org/10.1371/journal.pone.0099312
Martinez-Vicente, V., Evers-King, H., Roy, S., Kostadinov, T. S., Tarran, G. A., Graff, J. R., Brewin, R. J. W., Dall'Olmo, G., Jackson, T., Hickman, A. E., Röttgers, R., Krasemann, H., Marañón, E., Platt, T., & Sathyendranath, S. (2017). Intercomparison of ocean color algorithms for picophytoplankton carbon in the ocean. Frontiers in Marine Science, 4. https://doi.org/10.3389/fmars.2017.00378

Martinez Vicente, V., Sullivan, E., Harding, J., Atkinson, A., Raitsos, D., and Darmaraki, S. (2025). D4 Science Report BOOMS, Zenodo, https://doi.org/10.5281/zenodo.20670423

Menden-Deuer, S., & Lessard, E. J. (2000). Carbon to volume relationships for dinoflagellates, diatoms, and other protist plankton. Limnology and Oceanography, 45(3), 569–579. https://doi.org/10.4319/lo.2000.45.3.0569

Mie, G. (1908). Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. Annalen Der Physik, 330(3), 377–445. https://doi.org/10.1002/andp.19083300302

Moore, T., Campbell, J., & Feng, H. (2001). A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. IEEE Transactions on Geoscience and Remote Sensing, 39(8), 1764–1776. https://doi.org/10.1109/36.942555

Moore, T. S., Dowell, M. D., Bradt, S., & Verdu, A. R. (2014). An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. Remote Sensing of Environment, 143, 97–111. https://doi.org/10.1016/j.rse.2013.11.021

Morel, A., & Bricaud, A. (1986). Inherent properties of algal cells including picoplankton: theoretical and experimental results, Canadian Bulletin of Fisheries and Aquatic Sciences, 214, 521–559.

Morel, A., & Maritorena, S. (2001). Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research Atmospheres, 106(C4), 7163–7180. https://doi.org/10.1029/2000jc000319

Morel, A., Claustre, H., & Gentili, B. (2010). The most oligotrophic subtropical zones of the global ocean: similarities and differences in terms of chlorophyll and yellow substance. Biogeosciences, 7(10), 3139–3151. https://doi.org/10.5194/bg-7-3139-2010

NOAA National Geophysical Data Center (2009): ETOPO1 1 Arc-Minute Global Relief Model. NOAA National Centers for Environmental Information. Accessed 12/06/2026.

Opitz, D., & Maclin, R. (1999). Popular Ensemble Methods: an Empirical study. Journal of Artificial Intelligence Research, 11, 169–198. https://doi.org/10.1613/jair.614

O'Reilly, J. E., Maritorena, S., Siegel, D., O'Brien, M. C. (2000). Ocean color chlorophyll a algorithms for SeaWiFs, OC2, and OC4: Technical report. In D. Toole, B. G. Mitchell, M. Kahru, F. P. Chavez, P. Strutton, G. Cota, S. B. Hooker, C. R. McClain, K. L. Carder, F. Muller-Karger, L. Harding, A. Magnuson, D. Phinney, G. F. Moore, J. Aiken, K. R. Arrigo, R. Letelier, M. Culver, S. B. Hooker, & E. R. Firestone (Eds.), SeaWiFS postlaunch calibration and validation analyses, Part 3. NASA, Goddard Space Flight Center, Greenbelt, Maryland. SeaWiFS Postlaunch Technical Report Series, Vol. 11. (pp. 9–23).

Platt, T., & Sathyendranath, S. (1988). Oceanic Primary Production: estimation by remote sensing at local and regional scales. Science, 241(4873), 1613–1620. https://doi.org/10.1126/science.241.4873.1613

Platt, T., & Sathyendranath, S. (1993). Estimators of primary production for interpretation of remotely sensed data on ocean color. Journal of Geophysical Research Atmospheres, 98(C8), 14561–14576. https://doi.org/10.1029/93jc01001

Robertson Lain, L. R., & Bernard, S. (2018). The Fundamental contribution of Phytoplankton spectral scattering to ocean colour: Implications for Satellite detection of Phytoplankton community structure. Applied Sciences, 8(12), 2681. https://doi.org/10.3390/app8122681

Roy, S., Sathyendranath, S., & Platt, T. (2017). Size-partitioned phytoplankton carbon and carbon-to-chlorophyll ratio from ocean colour by an absorption-based bio-optical algorithm. Remote Sensing of Environment, 194, 177–189. https://doi.org/10.1016/j.rse.2017.02.015

Sathyendranath, S., & Platt, T. (1989). Computation of aquatic primary production: Extended formalism to include effect of angular and spectral distribution of light. Limnology and Oceanography, 34(1), 188–198. https://doi.org/10.4319/lo.1989.34.1.0188

Sathyendranath, S., & Platt, T. (2007). Spectral effects in bio-optical control on the ocean system. Oceanologia, 49, 5–39.

Sathyendranath, S., Stuart, V., Nair, A., Oka, K., Nakane, T., Bouman, H., Forget, M., Maass, H., & Platt, T. (2009). Carbon-to-chlorophyll ratio and growth rate of phytoplankton in the sea. Marine Ecology Progress Series, 383, 73–84. https://doi.org/10.3354/meps07998

Sathyendranath, S., Platt, T., Kovač, Ž., Dingle, J., Jackson, T., Brewin, R. J. W., Franks, P., Marañón, E., Kulk, G., & Bouman, H. A. (2020). Reconciling models of primary production and photoacclimation [Invited]. Applied Optics, 59(10), C100. https://doi.org/10.1364/ao.386252

Sheldon, R. W., Prakash, A., & Sutcliffe, W. H. (1972). The size distribution of particles in the ocean. Limnology and Oceanography, 17(3), 327–340. https://doi.org/10.4319/lo.1972.17.3.0327

Stock, A., & Subramaniam, A. (2020). Accuracy of Empirical satellite algorithms for mapping phytoplankton diagnostic pigments in the Open Ocean: A Supervised Learning perspective. Frontiers in Marine Science, 7. https://doi.org/10.3389/fmars.2020.00599

Sun, X., Brewin, R. J., Sathyendranath, S., Dall'Olmo, G., Airs, R., Barlow, R., Bracher, A., Brotas, V., Kheireddine, M., Lamont, T., Marañón, E., Morán, X. a. G., Raitsos, D. E., Shen, F., & Tilstone, G. H. (2022). Coupling ecological concepts with an ocean-colour model: Phytoplankton size structure. Remote Sensing of Environment, 285, 113415. https://doi.org/10.1016/j.rse.2022.113415

Taylor, M. H., Losch, M., & Bracher, A. (2012). On the drivers of phytoplankton blooms in the Antarctic marginal ice zone: A modeling approach. Journal of Geophysical Research Oceans, 118(1), 63–75. https://doi.org/10.1029/2012jc008418

Twardowski, M. S., Boss, E., Macdonald, J. B., Pegau, W. S., Barnard, A. H., & Zaneveld, J. R. V. (2001). A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. Journal of Geophysical Research Atmospheres, 106(C7), 14129–14142. https://doi.org/10.1029/2000jc000404

Uitz, J., Claustre, H., Morel, A., & Hooker, S. B. (2006). Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. Journal of Geophysical Research Atmospheres, 111(C8). https://doi.org/10.1029/2005jc003207

Van de Hulst, H. C. (1981). Light scattering by small particles, Dover Publications, New York, ISBN 10 0486642283, ISBN 13 9780486642284.

Vidussi, F., Claustre, H., Manca, B. B., Luchetta, A., & Marty, J. (2001). Phytoplankton pigment distribution in relation to upper thermocline circulation in the eastern Mediterranean Sea during winter. Journal of Geophysical Research Atmospheres, 106(C9), 19939–19956. https://doi.org/10.1029/1999jc000308

Volta, C., Bellacicco, M., Xing, X., Yang, G., & Marullo, S. (2025). Phytoplankton dynamics in subtropical gyres: New insights into biomass and physiology from 25 years of satellite observations. Geophysical Research Letters, 52(12). https://doi.org/10.1029/2024gl111817

Xi, H., Losa, S. N., Mangin, A., Soppa, M. A., Garnesson, P., Demaria, J., Liu, Y., D'Andon, O. H. F., & Bracher, A. (2020). Global retrieval of phytoplankton functional types based on empirical orthogonal functions using CMEMS GlobColour merged products and further extension to OLCI data. Remote Sensing of Environment, 240, 111704. https://doi.org/10.1016/j.rse.2020.111704

Xi, H., Losa, S. N., Mangin, A., Garnesson, P., Bretagnon, M., Demaria, J., Soppa, M. A., D'Andon, O. H. F., & Bracher, A. (2021). Global Chlorophyll A concentrations of phytoplankton functional types with detailed uncertainty assessment using multisensor ocean color and sea surface temperature satellite products. Journal of Geophysical Research Oceans, 126(5). https://doi.org/10.1029/2020jc017127

Xi, H., Bretagnon, M., Losa, S. N., Brotas, V., Gomes, M., Peeken, I., Alvarado, L. M. A., Mangin, A., & Bracher, A. (2023). Satellite monitoring of surface phytoplankton functional types in the Atlantic Ocean over 20 years (2002–2021). State of the Planet, 1-osr7, 1–13. https://doi.org/10.5194/sp-1-osr7-5-2023

Xi, H., Bretagnon, M., Mehdipour, E., Demaria, J., Mangin, A., & Bracher, A. (2025). Consistent long-term observations of surface phytoplankton functional types from space. State of the Planet, 6-osr9, 1–15. https://doi.org/10.5194/sp-6-osr9-7-2025

Yang, G., Bellacicco, M., Organelli, E., & Xing, X. (2024). Global variability of phytoplankton carbon and Non-Algal particles from ocean color data based on a photoacclimation model. Journal of Geophysical Research Oceans, 129(1). https://doi.org/10.1029/2023jc019922

Yang, X., Sun, H., Fu, K., Yang, J., Sun, X., Yan, M., & Guo, Z. (2018). Automatic Ship Detection in Remote Sensing Images from Google Earth of Complex Scenes Based on Multiscale Rotation Dense Feature Pyramid Networks. Remote Sensing, 10(1), 132. https://doi.org/10.3390/rs10010132

Zhang, X., Lewis, M., Lee, M., Johnson, B., & Korotaev, G. (2002). The volume scattering function of natural bubble populations. Limnology and Oceanography, 47(5), 1273–1282. https://doi.org/10.4319/lo.2002.47.5.1273