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

ESA 4000147645/25/I-LR
D3.1 System Specification Document (SSD)

Project led by

Plymouth Marine Laboratory

In cooperation with

Document properties

Project:	PHYTO-CCI
Document title:	System Specification Document (SSD)
Reference:	D3.1
Date:	19/06/2026
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 11:48:06 +01'00'

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

Copyright: © Plymouth Marine Laboratory 2026. The copyright in this document is vested in Plymouth Marine Laboratory. This document may only be reproduced in whole or in part, or stored in a retrieval system, or transmitted in any form, or by any means electronic, mechanical, photocopying or otherwise, either with the prior permission of Plymouth marine Laboratory or in accordance with the terms of ESA Contract No. 4000147645/25/I-LR.

Document version history

Version	Date	Author(s) / Contributor(s)	Comment
0.0	18/12/2025	Gemma Kulk	Template
1.0 (draft)	18/06/2026	Gemma Kulk	First draft
1.0	19/06/2026	Gemma Kulk, Salem I. Salem	First release

Index

1 Introduction

1.1 Purpose of document

This System Specification Document (SSD) described the processing system of the PHYTOplankton biomass and diversity Climate Change Initiative (PHYTO-CCI). The system is based on that of the Ocean Colour Climate Change Initiative (OC-CCI), ingesting remote sensing reflectances, inherent optical properties and chlorophyll-a to derive Essential Climate Variables (ECVs) for phytoplankton carbon biomass and pigment diversity. The SSD is written with the intention to cover the first development round in PHYTO-CCI, with further developments of the processing system planned in consecutive development rounds. It should therefore be seen as a living document until the final ocean biosphere ECV climate data record is produced and made publicly available.

1.2 Structure

The SSD summaries the processing system of OC-CCI, that forms the basis of the processing system of PHYTO-CCI, in Section 2. The current version of the processing system of PHYTO-CCI is described in detail in Section 3, including requirements, design, data access and future developments. A summary is provided in Section 4 and references in Section 5.

1.3 Related documents

Documents related to the PHYTO-CCI SSD are the PHYTO-CCI User Requirement Document (URD), Product Specification Document (PSD), Product Validation and Algorithm Selection Report (PVASR), Algorithm Theoretical Baseline Document (ATBD), End-to-End ECV Uncertainty Budget (E3UB), Algorithm Development Plan (ADP) and the Product Validation Plan (PVP).

2 OC-CCI processing system

While the phytoplankton carbon and pigment products are intended as a new ocean biosphere ECV, the products are based on ocean-colour observations and are therefore closely related to the ocean-colour ECV produced by the OC-CCI. We therefore plan to use the processing system of the OC-CCI as a starting point, for consistency between ocean-colour-derived ECV products, and for which we can also rely on the expertise and resources available within OC-CCI. Below we provide a brief description of the OC-CCI processing chain. Detailed information on the OC-CCI processing chain, including system design and infrastructure, processing workflows and performance is provided in the OC-CCI System Specification Document (available at: https://docs.pml.space/share/s/6wGtN6XUR_-37HiOXXlZvg) and the annex to this document the Calvalus Technical Specification (available at: https://docs.pml.space/share/s/5-so3GbvTEusu3mOWCOQAw).

2.1 Summary

The current version of the OC-CCI time series (version 6) incorporates ocean colour observations from six different satellite sensors, including Sea-Viewing Wide Field-of-View Sensor (SeaWiFS), Moderate Resolution Imaging Spectroradiometer (MODIS), Envisat MEdium Resolution Imaging Spectrometer (MERIS), Visible Infrared Imaging Radiometer Suite (VIIRS), and Sentinel-3 Ocean and Land Colour Instrument (OLCI) A and B, with MERIS acting as the reference sensor. In Table 1 we detail the specific input datasets and the period covered, as well as the approximate volume of data for each individual sensor.

*Table 1. Overview of ocean colour datasets included in the processing chain of OC-CCI version 6. * Note that the addition of NOAA-20 VIIRS is planned for the next phase of OC-CCI.*

Dataset	Period covered	Approximate volume
SeaWiFS L1 GAC & LAC	1997-2010	5 TB
SeaWiFS L2 GAC & LAC	1997-2010	5 TB
MERIS 1-km L1b 4th reprocessing	2002-2012	25 TB
MODIS-AQUA L1A	2002-2019	250 TB
MODIS-AQUA L2	2002-2019	40 TB
Suomi VIIRS L1A	2011-2019	300 TB
NOAA-20 VIIRS L1A*	2017-present	200 TB
OLCI-3A 1-km	2016-present	180 TB per year
OLCI-3B 1-km	2018-present	180 TB per year

As a first step in the OC-CCI processing chain, an atmospheric correction algorithm is applied to data from individual sensors to obtain level 2 data. MODIS, VIIRS, MERIS and OLCI data are processed with the POLYMER algorithm (v4.14), while SeaWiFS level 2 data are downloaded from National Aeronautics and Space Administration (NASA), implicitly meaning that these data are processed with the l2gen algorithm from SeaDAS (v9.7.0-V2024.6). Data is flagged and masked using schemes from multiple sources, including POLYMER and IDEPIX (v6.0) bitmasks and IDEPIX (v6.0/7.0) cloud and land flags as well as standard NASA level 2 flags. Next, all individual sensors are binned using the BEAM/SNAP binner to obtain level 3 data at 4 km resolution on a sinusoidal grid. The obtained daily remote sensing reflectances (R_{rs}) for SeaWiFS, MODIS, VIIRS and OLCI are then band-shifted to the six main bands of the reference sensor MERIS (412, 443, 490, 510, 560, 665 nm) by computing quasi-analytical algorithm (Mélin-QAA) inherent optical properties (IOPs) and back computing the R_{rs} bands using a high-resolution spectral model. Negative values in the output R_{rs} are either removed (for bands between 412-560 nm) or clamped to zero (for 665 nm). Note that no band-shifting is done to the MERIS data in this step. The fourth step addresses biases between individual sensors. A pixel-by-pixel bias correction for R_{rs} from SeaWiFS and MODIS is performed in respect to MERIS using a temporally-weighted climatology (7 days), such that the corrections take account of seasonal and regional variations. The biases for SeaWiFS and MODIS are computed for 2003-2007 as all sensors required (including MERIS) are overlapping and functioning well. Bias adjustments are computed at every location where all sensors gathered data, with a temporal window of +/- 45 days and spatially-limited interpolation (11 pixels) to fill smaller gaps. VIIRS and OLCI-A are bias corrected to MERIS using a similar process, but comparing against MODIS-corrected-to-MERIS-levels rather than directly to MERIS. This indirect comparison is unavoidable due to the lack of temporal overlap among MERIS and OLCI, and made use of data spanning 2012-2014 for VIIRS and 2016-2019 for OLCI-A. Similarly, OLCI-B lacks overlap with MERIS and is corrected to the 'MERIS-corrected' OLCI-A using data from 2018-2021. Following this step of de-biasing, the individual sensor R_{rs} were merged with a simple average. The multi-sensor R_{rs}, i.e., the band-shifted, bias-corrected and merged R_{rs}, are then used to compute optical water classes according to Jackson *et al.* (2017) to assign Fuzzy C-Means clustering membership of 14 classes to each pixel. A range of products were then generated from the merged R_{rs} using the validated algorithms in SeaDAS, including IOPs based on the QAA and a merged chlorophyll-a product based on the OCI, OCI2, OC2 and OCx algorithms, weighted by the relatively level of membership in specific optical water classes. $K_d(490)$ (Lee variant) is generated independently due to implementation issues in the SeaDAS variant. A table of uncertainties for each optical water class is computed from match-up *in situ* and the produced OC-CCI data, from which pixel-specific total uncertainty is computed using the optical water class membership for each pixel. Final steps in the OC-CCI processing chain include reprojection onto a geographic grid (in addition to the basic sinusoidal grid), production of 5-day, 8-day, monthly and annual composites and production of data at lower temporal resolutions (for example at 0.25 and 0.5 degree upon request for Obs4MIPs). In addition, PNG quicklooks are created for all products following scaling factors from NASA.

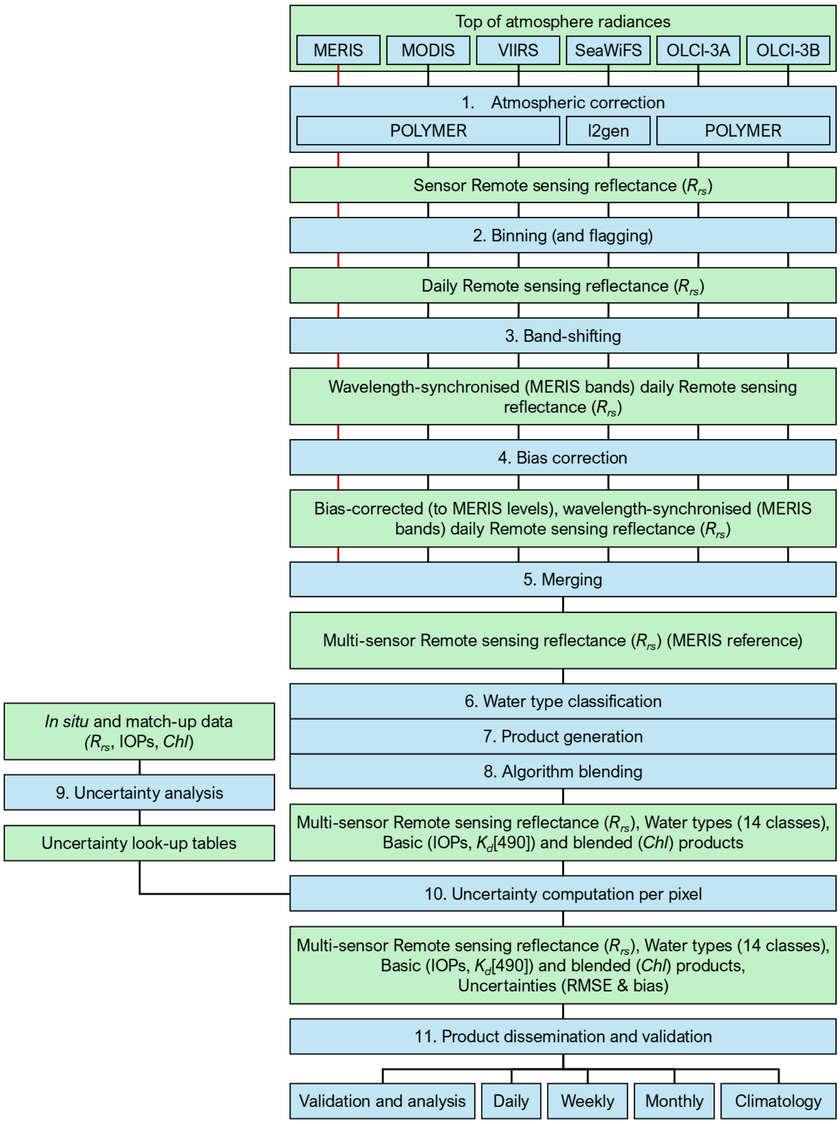

Figure 1. Date flow in the Ocean Colour ECV production of the Ocean Colour Climate Change Initiative (OC-CCI). Figure from OC-CCI version 6 user manual (Sathyendranath, 2022).

3 PHYTO-CCI processing system

3.1 Requirements

The user consultation of PHYTO-CCI identified requirements of various user groups on the spatial and temporal resolution of the ocean biosphere ECV products, but also on data access and services (see for details the User Requirements Document (URD)). For the development of the first ECV dataset of phytoplankton carbon biomass and phytoplankton pigment diversity, several algorithms were selected and data was produced at 4 km spatial and monthly temporal resolution, with a regional test dataset at 5-day resolution for the pigment fucoxanthin. While this resolution is close to the threshold level identified by users, higher resolutions were identified at goal level with preferences for ECV data at 0.1-1 km spatial resolution and 1-day temporal resolution (Table 2).

Table 2. Summary of requirements for Essential Climate Variables (ECVs) and associated products related to phytoplankton biomass and diversity according to the online user consultation survey. Requirements for the spatial and temporal resolution are provided at three levels of maturity: the ideal requirement above which no further improvements are needed (G – Goal), an intermediate level at which significant improvements are made (B – Breakthrough), and the minimum requirements to ensure usefulness of the data product (T – Threshold). Adapted from Table 4 in User Requirements Document (URD).

	ECV	Plankton	
	Product	Phytoplankton carbon	Phytoplankton pigments
Requirement	Level	Values	Values
Horizontal resolution	G B T	0.1-1 km 0.1-1 km 1-10 km	0.1-1 km 0.1-1 km 1-10 km
Temporal resolution	G B T	1 day 1 day 1 week	1 day 1 day 1 week

It is noted that the satellite-retrieval algorithms are primarily based on input variables from OC-CCI and therefore the spatial and temporal resolution of the ocean biosphere ECV is dictated by the resolution available for this input data. For example, the spatial resolution identified by users at goal level, i.e., 0.1-1 km, would be difficult to achieve based on the current merged ocean-colour time series from OC-CCI. While chlorophyll-a is produced at 1 km resolution, other ocean-colour variables are currently available at 4 km resolution. This is not a limitation set by OC-CCI, but rather a limitation of the original ocean-colour sensors, with only Sentinel-3 data available at higher resolutions (300 m). The same applies to satellite-based data from other sources, such as Sea Surface Temperature Climate Change Initiative (SST-CCI), which is currently available at 5 km spatial resolution. The input data to the PHYTO-CCI processing system are further detailed in Section 3.2.3.

The technical requirements identified by users included a preference for data in netCDF and geoTIFF formats and access to data via a web browser, file transfer protocol (FTP) server or cloud, either through services provided by Plymouth Marine Laboratory (i.e., OC-CCI open data portal) or Copernicus Climate Change Service (C3S) (Figure 2). All these requirements should be considered in the dissemination of the ocean biosphere ECV data products, keeping in mind that users are typically able to use multiple file formats and data services.

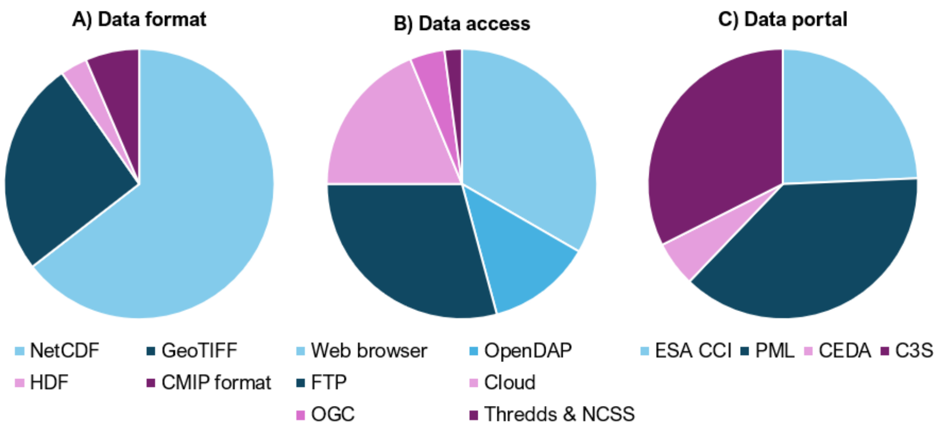

Figure 2. Data level, format and access, as polled among participants of the online user requirements survey, based on their experiences with OC-CCI and other ocean-colour data. Abbreviations: NetCDF – Network Common Data Form, HDF – Hierarchical Data Format, GeoTIFF – Georeferenced Tagged Image File Format, CMIP – Coupled Model Intercomparison Project, FTP – File Transfer Protocol, OGC – Open Geospatial Consortium, OpenDAP – Open-source Project for a Network Data Access Protocol, NCSS – NetCDF Subset Service, ESA – European Space Agency, CCI – Climate Change Initiative, PML – Plymouth Marine Laboratory, CEDA – Centre for Environmental Data Analysis, C3S – Copernicus Climate Change Service. From User Requirements Document (URD).

The system requirements are designed to accommodate the user requirements, but also any technical requirements that are needed for algorithm development, intercomparison and validation, and ECV data production and dissemination. The requirements identified for algorithm development, intercomparison and validation as well as ECV data production include:

- Storage of data, including input, intermediate, output and auxiliary data
- Ingestion of new data sources, specifically SST-CCI
- Data processors to transform input into outputs
- Control of processing workflows, in particular for parallel (bulk) data production
- Data quality checks, both automated and visual
- Comparison with reference data through match-up and validation
- Documentation, including metadata, to distinguish versions

The requirements identified for dissemination of data products and documentation include:

- Data access online, including (bulk) access to complete data records
- Tools for data customisation
- Documentation on products, including user guide with examples
- Feedback handling and issue tracking
- Long-term preservation of the ECV data record

While many of these user and system requirements are met by the OC-CCI processing system, there is further system development needed as part of the PHYTO-CCI to be able to produce the ocean biosphere ECV products. Initial steps have been undertaken in the first development round of PHYTO-CCI, but more work is needed for the production and dissemination of the final data record in the next development rounds.

3.2 Design

3.2.1 Functionality

The design of the ECV processing system is important to consider for efficiency, traceability, versioning and updating as well as transitioning into an operational service. The system should be designed around the users of the climate research community, who require state-of-the-art time-series products that are both periodically updated to incorporate advances in methodology and new data, and completely reprocessed in a systematic and consistent manner over the entire time span. This requires an agile, responsive and efficient processing chain, capable of incorporating new developments in a modular fashion. The basics of such a system have been developed in OC-CCI and we use the functionalities in their processing chain as a starting point for the production of the targeted ocean biosphere ECV products. While the functionalities meet the system requirements as identified in Section 2.1 for the variables produced by OC-CCI, we note that further development is needed to incorporate the specific components of the PHYTO-CCI processing chain to produce the ocean biosphere ECV data record.

3.2.2 Architecture

The processing system of OC-CCI until step 10 (Figure 1), which produces multi-sensor remote sensing reflectance, optical water types, inherent optical properties and blended chlorophyll-a products with their uncertainties, forms the starting point of the processing system for the ocean biosphere ECV (Figure 3). The PHYTO-CCI processing system starts with the generation of the phytoplankton carbon and pigment data from individual algorithms using as input the OC-CCI data and other EO data, including other CCI products, and re-analysis data as required by the selected algorithms (see Section 3.2.3). Depending on the outcome of the algorithm intercomparison, the data is then blended according to water class membership, following the approach described by Jackson *et al.* (2017). A match-up dataset of *in situ* data and individual or blended phytoplankton carbon and pigment data from PHYTO-CCI is then used to produce look-up tables to assign uncertainties per-pixel depending on membership of optical water classes of each pixel. The multi-sensor (merged) phytoplankton carbon and pigment products and their uncertainties are then prepared for validation and dissemination at appropriate temporal and spatial resolutions. The processing system will also automate the production of data for validation and generate the required match-up of independent *in situ* data with the individual or blended ECV products.

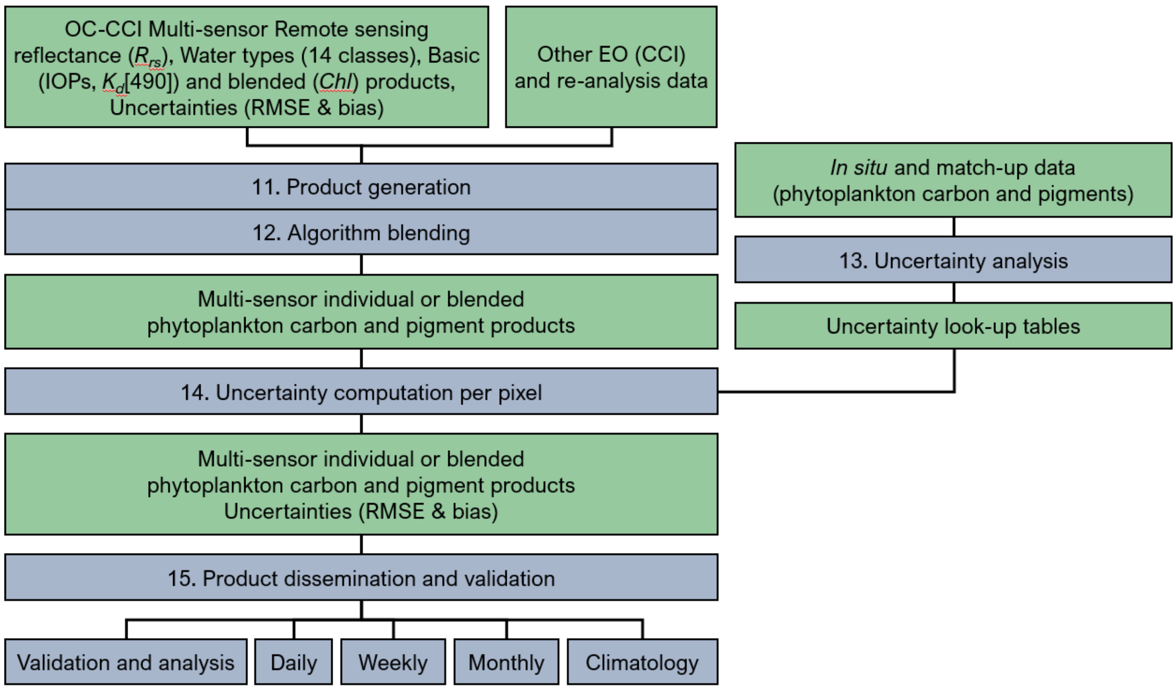

Figure 3. Design of the PHYTO-CCI processing system, which is based on the OC-CCI processing system. The processing system of OC-CCI until step 10 (Figure 1), which produces multi-sensor remote sensing reflectance (R_{rs}*), optical water types, inherent optical properties (IOPs) and blended chlorophyll-a (*Chl*) products with their uncertainties, forms the basis of the processing system for the phytoplankton biomass and diversity ECV products. The PHYTO-CCI processing system then generates the products for individual algorithms and/or blended algorithms with input from OC-CCI products and other EO (including other SST-CCI) and match-up* in situ *data of phytoplankton carbon and pigments. Multi-sensor, individual and/or blended phytoplankton carbon and pigments products and their uncertainties are then disseminated for validation and to the user community at the required temporal and spatial resolutions.*

In the first development round of PHYTO-CCI, several components of the processing system have been developed that are modular and can be added to the processing system of OC-CCI. These developments include collation, quality control and match-up of *in situ* databases of phytoplankton carbon and pigments, to facilitate algorithm development, intercomparison and validation. The quantitative scoring system used in the algorithm intercomparison was updated for phytoplankton carbon and pigment data. In addition, several of the phytoplankton carbon and pigment algorithms as well as uncertainty estimation based on optical water types have also implemented for production of the ECV data record at 4 km spatial and monthly temporal resolution (Table 3). We note that some data products were produced externally, by individual algorithm developers, and this will be reviewed in the second development round of PHYTO-CCI. In the first development round, preparations have also been made to produce data at additional spectral bands, which are needed for the improvement of phytoplankton pigment algorithms. This requires significant changes to the underlying processing chain within OC-CCI, which currently produces data at six wavelengths.

Table 3. Implementation of new algorithms for production of the ocean biosphere ECV in the PHYTO-CCI processing system.

Algorithm	Underlying method	Implemented in PHYTO-CCI system	Processing times	Key reference(s)
Phytoplankton carbon biomass				
A	Particle back-scattering-based empirical algorithms	Yes	Day	Bellacicco *et al.,* 2020; Yang *et al.*, 2024
B_1 (tuned) B_2 (untuned)	Particle-backscattering-based allometric semi-analytical models	No, data produced by member of advisory group	Weeks	Kostadinov *et al.*, 2023
C	Chlorophyll-a-based empirical algorithms	Yes	Day	Marañón *et al.*, 2014
D	Photoacclimation-based models	Yes	Month	Sathyendranath *et al.*, 2020
Phytoplankton pigment diversity				
A_1 (without SST) A_2 (with SST)	Spectral decomposition & generalised linear model fitting	No, data produced by project partner	Week	Bracher *et al.* (2015); Xi *et al.* (2020)
B_1 (without SST) B_2 (with SST)	Deep learning model with fusion architecture	No, data produced by member of advisory group	Week	Li *et al.* (2023)
C	Random Forest model	Yes	Days	Laurenson *et al.* (in preparation)

3.2.3 Inputs

The primary input to the phytoplankton carbon and pigment algorithms for production of the ocean biosphere ECV are observations of OC-CCI, including the remote sensing reflectances at 412, 443, 490, 510, 560 and 665 nm, the particle backscattering coefficient at 443 nm and chlorophyll-a (Table 4). These products are based on several ocean-colour sensors as described in Section 2 (also see Table 1). Additional satellite-based variables are needed from other sources for implementation of specific phytoplankton carbon and pigment algorithms (Table 4; see for details the Algorithm Theoretical Baseline Document (ATBD)). For one phytoplankton carbon algorithm, information on Photosynthetically Active Radiation (PAR) and mixed layer depth is needed. These variables are currently not available as climate data records from CCI, and instead single sensor products of PAR from NASA are bias-corrected and merged to form a consistent time series, and a climatology of mixed layer depth is used (De Boyer Montégut *et al.*, 2004), which is based on *in situ* observations. For the same phytoplankton carbon algorithm, maps of model parameters are needed, which are based on *in situ* observations and assigned per ecological province and per season (Kulk *et al.*, 2020). For another phytoplankton carbon algorithm, a climatology of the backscattering of non-algal pigments is needed, which is derived from a combination of satellite observations and re-analysis datasets (Volta *et al.*, 2025). For two phytoplankton pigment algorithms, SST from CCI is used as an additional input, and for the other phytoplankton pigment algorithm, information on bathymetry is needed. These additional satellite-based and other mapped data products are specifically incorporated into the processing system for the production of the ocean biosphere ECV in PHYTO-CCI. It is noted that the input data is available at various spatial resolutions, and there is a need to reproject the data to the same grid as that of OC-CCI data, which is an functionality in the processing system.

Table 4. Overview of the input data required for implementation of the individual phytoplankton carbon (PC) and phytoplankton pigment (PG) algorithms, with information on data sources and highest resolution available from those sources. Note that for some products, reprojection to the OC-CCI grid is needed for implementation of the algorithms at 4 km resolution. Details of the input requirements for each algorithm are detailed in the ATBD.

Input variable	**Source**	**Highest resolution available**	**Required for algorithm(s)**
Remote sensing reflectances	OC-CCI v6	4 km, daily	PC: B_1, B_2 PG: A_1, A_2, B_1, B_2, C
Particle backscattering coefficient	OC-CCI v6	4 km, daily	PC: A
Chlorophyll-a	OC-CCI v6	1 km, daily	PC: A, C, D
Sea Surface Temperature	SST-CCI v3	5 km, daily	PG: A_2, B_2
Photosynthetically Active Radiation	NASA R2022.0	4 km, daily	PC: D
Mixed layer depth climatology	De Boyer Montégut *et al.*, 2004	2 °, monthly	PC: D
Bathymetry	GEBCO v2026	4 km, monthly	PC: D PG: C
Photosynthetic parameters	Kulk *et al.*, 2020	4 km, seasonally	PC: D
Chlorophyll-a profile parameters	Kulk *et al.*, 2020	4 km, seasonally	PC: D
Backscattering of non-algal particles climatology	Volta *et al.*, 2025	25 km, annually	PC: A

For algorithm development, intercomparison and validation, *in situ* observations of phytoplankton carbon and pigments are needed. PHYTO-CCI collated *in situ* database for the ocean biosphere ECVs, with observations of phytoplankton carbon based on flow cytometry measurements and observations of phytoplankton pigments based on High Performance Liquid Chromatography (HPLC). The details of the *in situ* datasets are provided in the Product Validation and Algorithm Selection Report (PVASR) and the Product Validation Plan (PVP). The *in situ* data are also used to establish uncertainty look-up tables to assign per-pixel bias and RMSD for each phytoplankton carbon and pigment algorithm (see for details End-to-End ECV Uncertainty Budget (E3UB)).

3.2.4 Outputs

The PHYTO-CCI processing system produces the following ocean-colour products for the ocean biosphere ECV:

- *Phytoplankton biomass* as the weight of carbon per volume, i.e., phytoplankton carbon in mg m^{-3}. Initially, the total phytoplankton carbon concentration is produced, where possible, the carbon content of three phytoplankton size classes, i.e., pico-, nano- and microphytoplankton will be produced in later development rounds.

- *Phytoplankton diversity* will be provided as the concentration of pigment types per volume, i.e., specific phytoplankton pigments in mg m^{-3}. Eight phytoplankton pigments are produced: 19'-butanoyloxyfucoxanthin, 19'-hexanoyloxyfucoxanthin, alloxanthin, divinyl chlorophyll-a, chlorophyll-b, fucoxanthin, peridinin and zeaxanthin.

Note that chlorophyll-a, the main photosynthetic pigment in phytoplankton, and also a measure of phytoplankton biomass, will not be produced in PHYTO-CCI, as it is already available through OC-CCI. An overview of the ocean biosphere ECV data products is provided in Table 5 and further details are specified in the Product Specification Document (PSD). In the first development round, phytoplankton carbon products from four different algorithms (Algorithms A, B_1, C and D) and phytoplankton pigment products from the best-performing algorithm (Algorithm B_2) are produced at 4 km spatial and monthly temporal resolution. All products are produced with per-pixel uncertainties based on an optical water type approach (see E3UB). The data are stored in netCDF file format.

Table 5. Overview of the first version of the ocean biosphere ECV with the core ECV products phytoplankton carbon and pigments. The included sensors, spatial and temporal coverage and resolution and data size are provided with a brief explanatory text for each product.

Product	Sensors	Spatial coverage	Spatial resolution	Temporal coverage	Temporal resolution	Data size	Explanatory text
Phytoplankton carbon (total phytoplankton carbon) [mg m^{-3}] GCOS ECV	SeaWiFS MERIS MODIS VIIRS OLCI-A OLCI-B	Global	4x4 km	1997-2027	Monthly	4 TB	Level 3-4 binned, multi-sensor merged, monthly products with uncertainty. Four products are available, based on four individual algorithms. This is one of the main PHYTO-CCI ECV products.
Phytoplankton pigments (19'-butanoyl-oxyfucoxanthin, 19'-hexanoyloxifucoxanthin, alloxanthin, chlorophyll-b, divinyl-chlorophyll-a, fucoxanthin, peridinin, zeaxanthin) [mg m^{-3}] GCOS ECV	SeaWiFS MERIS MODIS VIIRS OLCI-A OLCI-B	Global	4x4 km	1997-2027	Monthly	4 TB	Level 3-4 binned, multi-sensor merged, monthly with uncertainty. Based on best-performing algorithm. This is one of the main PHYTO-CCI ECV products.
Phytoplankton pigments (fucoxanthin) [mg m^{-3}] GCOS ECV	SeaWiFS MERIS MODIS VIIRS OLCI-A OLCI-B	Regional	4x4km	1997-2027	5-day	<1 TB	Level 3-4 binned, multi-sensor merged, 5-day with uncertainty. Based on best-performing algorithm. This is an experimental product.

3.2.5 Verification

Verification withing the PHYTO-CCI processing chain focusses on the demonstration that the system meets the specified requirements and provides the correct data to users. Verification methods follow those of OC-CCI and for the first development round, we used testing, monitoring and inspecting techniques:

- Additional code developed for the processing system for PHYTO-CCI were tested during software development and before deployment to make sure that the results are as expected
- Completeness of the data production was monitored by inspecting report files and checking that the expected output files were generated
- A random selection of data files were visually inspected, by opening files with appropriate software and looking at the results.

In the next development rounds, further verification will be carried out to make sure that results are the same after software changes, primarily targeted at optimising algorithm code.

3.2.6 Data access

The first version of the ocean biosphere ECV is intended for use by the Climate Research Group, which form part of the PHYTO-CCI team either as project partners or external collaborators. As this is a relatively small group of users, the first version of the dataset will not be made publicly available, also considering that improvement of phytoplankton carbon and pigments algorithms, and hence the ECV dataset, is foreseen in the second and third development rounds. Data access will therefore initially be arranged through an ftp server, from which the netCDF files can be downloaded.

The final version of the ocean biosphere ECV dataset, i.e., after the third development round, will be made publicly available. As a first step, the ocean biosphere ECV data record will be archived at CEDA, for which we already identified data storage requirements and requested access to the archive. This will be linked to the CCI Open Data Portal, which gives users a direct link to download data from the archive. Then, data will be made publicly available through an ftp server, which is a preferred choice for many users, and we will make use of the services provided by OC-CCI. Finally, the phytoplankton carbon and pigment products will be added to the OC-CCI data portal, which is a web GIS portal that allows viewing, manipulating and analysis of data, extending the potential user base considerably.

3.2.7 Source code and version control

The software used by PHYTO-CCI is partly based on existing software within the processing system of OC-CCI, which is under configuration control and managed by the development team. Where needed, PHYTO-CCI team members have access to this code base that is version controlled using git. The code specifically developed for PHYTO-CCI is written in Python 3 and version controlled using GitLab.

3.3 Future developments

In the second development round of PHYTO-CCI, we foresee improvements of existing components and addition of new components to the processing system. These include:

- Implementation of code of all phytoplankton carbon and pigment algorithms. As described in Section 3.2.2, some of individual ECV datasets at 4km, monthly resolution were produced externally and there is a need to implement the code of these algorithms in the processing system, to streamline workflows as well as ensure consistency in data production.
- The production of data from one phytoplankton carbon algorithm requires significant computational resources and time (Algorithm C), and it should be considered how this algorithm could be optimised. This is especially important for data production at higher resolution is to be considered in the next development rounds.
- There is a requirement for additional OC-CCI wavebands to improve the performance of the phytoplankton pigment algorithms (see ATBD and PVASR), which requires changes to the current processing system of OC-CCI and will be implemented as part of version 7.
- The implementation of a formal error propagation approach, using first order first moment (FOMF) or Monte Carlo methods, will also require changes to the current processing system of OC-CCI, and are likely to only be implemented for the ocean biosphere ECV products.
- Updates relating to keeping products state-of-the-art, utilising new algorithms and approaches for product generation and blending, that will emerge from the algorithm development tasks will also be addressed in consecutive cycles of ECV data production.
- Further discussion will be held with the OC-CCI development team to ensure that the final ocean biosphere ECV data record can be disseminated through their services (see Section 3.2.6).

As part of the above development points, cost and performance are considered to make the system as efficient as possible. The OC-CCI team has experience in this and we will collaborate with them to achieve an efficient processing system that can produce all ocean-colour derived variables.

4 Summary

The PHYTO-CCI processing system has a foundation in the OC-CCI processing system, from which data is ingested to produce the ocean biosphere ECV data products. The processing system uses the input data to initially produce data products from individual algorithms, at the preferred spatial and temporal resolution. An optical water type approach is then be used to merge data products (depending on the outcome of the algorithm intercomparison) and assign uncertainties, either to the individual data product or the merged data product. The individual or merged data products, including uncertainties, are then disseminated through appropriate channels. The phytoplankton carbon biomass and phytoplankton pigment diversity data products are primarily based on ocean colour observations from SeaWiFS, MERIS, MODIS, VIIRS and OLCI, which are merged in one time series by OC-CCI, and on observations from SST-CCI. The system produces the ocean biosphere ECV data record as the main output, with the first version consisting of phytoplankton carbon products from four algorithms and phytoplankton pigment products from one best-performing algorithm, available at 4 km spatial and monthly temporal resolution between September 1997 and December 2025, in netCDF file format.

In the first development round of PHYTO-CCI, several components of the processing system have been developed that are modular and can be added to the processing system of OC-CCI. These developments include collation, quality control and match-up of *in situ* databases of phytoplankton carbon and pigments, to facilitate algorithm development, intercomparison and validation. The quantitative scoring system used in the algorithm intercomparison was updated for phytoplankton carbon and pigment data. In addition, several of the phytoplankton carbon and pigment algorithms as well as uncertainty estimation based on optical water types have also implemented for production of the ECV data record. Verification processes have been put in place to assure that the processing system meets the requirements. Access to the data is provided initially through an ftp server to the Climate Research Group. The next developments rounds will focus on implementation and optimisation of the various satellite-retrieval algorithms needed for data production, implementation of a new approach to error propagation, production of data at additional spectral bands (in collaboration with OC-CCI) and preparation of dissemination of data and documentation for the final ECV dataset.

5 References

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

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

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

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

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

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

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

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

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

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