

D1.1 Protocol for pelagic habitats and FW2 data ingestion

1 Background

The OSPAR pelagic habitat monitoring programme is made up of multiple independent plankton monitoring programmes, including both continuous and station-based sampling designs (**Figure 1**). Each separate institution contributing data to pelagic habitats indicator assessment is responsible for managing their own dataset.

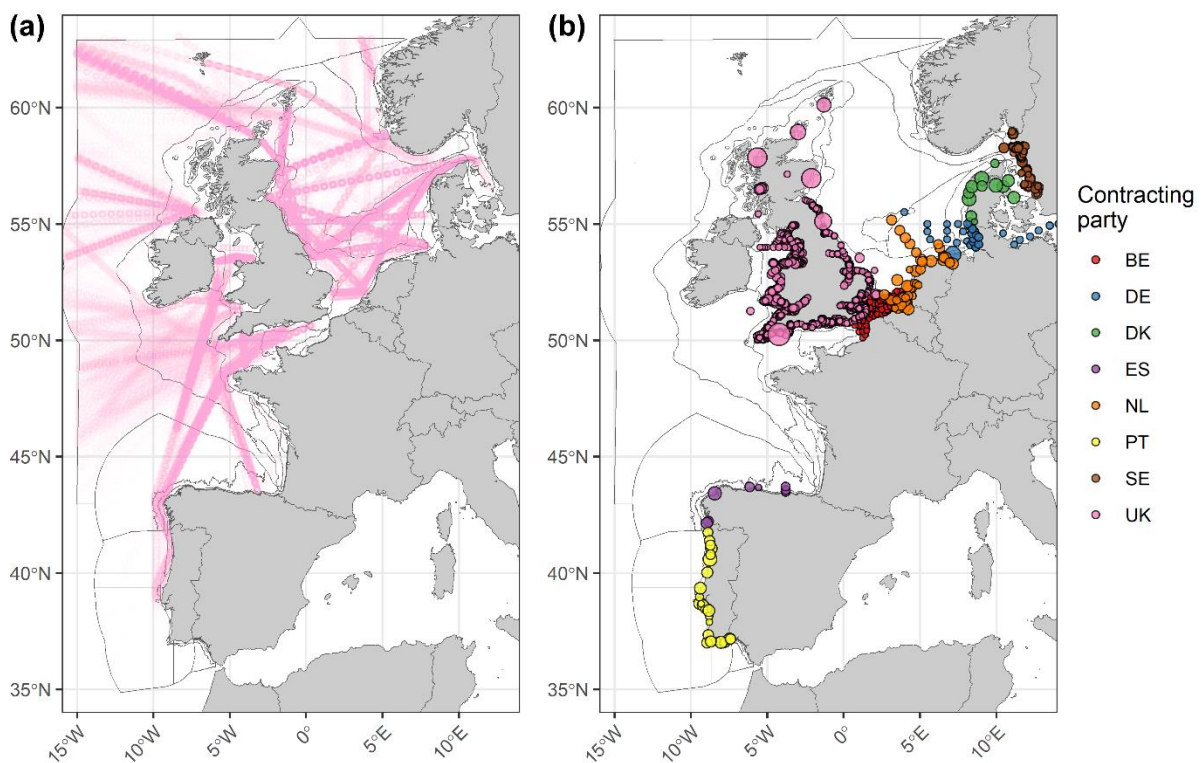


Figure 1. Locations of the Continuous Plankton Recorder (CPR) samples (a) and locations of the national plankton timeseries datasets (b) currently integrated in data flows for the Plankton Community Indicator. Datasets are grouped by contracting party.

To contribute to a better understanding of the state of pelagic habitats throughout the Greater North Sea, Celtic Seas, Bay of Biscay and Iberian Coast, Contracting Parties are required to report data (as specified below) that will enable an assessment of three common biodiversity Indicators and one candidate indicator:

- **PH1/FW5 – Change in plankton communities**, in OSPAR Region II - Greater North Sea, OSPAR Region III - Celtic Seas, and OSPAR Region IV - Bay of Biscay and Iberian Coast.
- **PH2 – Plankton abundance/biomass**, in OSPAR Region II - Greater North Sea, OSPAR Region III - Celtic Seas, and OSPAR Region IV - Bay of Biscay and Iberian Coast.
- **PH3 – Plankton diversity indices**, is common in OSPAR Region III – Celtic Seas and is a candidate indicator in OSPAR Region II - Greater North Sea and OSPAR Region IV Bay of Biscay and Iberian Coast.

- **FW2 – Productivity of phytoplankton**, is a pilot assessment in OSPAR Region II – Greater North Sea, OSPAR Region III – Celtic Seas and OSPAR Region IV – Bay of Biscay and Iberian Coast.

The data forms the basis for the above indicator assessments. Data are stored and made available via the *Data Archive for Species and Habitats* (DASSH; <https://www.dassh.ac.uk/>) and support delivery for future national and regional *Marine Strategy Framework Directive* (MSFD) assessments.

DASSH also host and manage the *Plankton Lifeform Extraction Tool* (PLET: <https://www.dassh.ac.uk/lifeforms/>), which ingests the raw species/taxa-level abundances and outputs aggregated monthly lifeform time-series data products through querying a Master Taxa List of functional trait information. The full functionality of the PLET is detailed in Ostle et al. (2021). This tool allows data providers to convert their datasets into the lifeforms required for PH1/FW5 and PH3, a normally resource-intensive process. The ongoing utility of the PLET and the Master Taxa List will be dependent on expert advice contributed by data providers as the knowledge base of taxonomic and functional information grows. The Master Taxa List can be downloaded from the PLET tool (<https://www.dassh.ac.uk/doitool/data/1709>), with each version assigned a unique Digital Object Identifier (DOI) for transparency of data processing.

For the OSPAR regional assessment, the geographic extent of the assessments in each Region are dependent on the distribution, amount, and quality of data supplied by Contracting Parties. The data are used to construct regional indicators. New data requests will serve to extend the time-series for current indicators and to collate all available data on plankton abundance, plankton biomass, and primary productivity for OSPAR Regions II, III, and IV to be made available for public access and use (**Figure 2**).

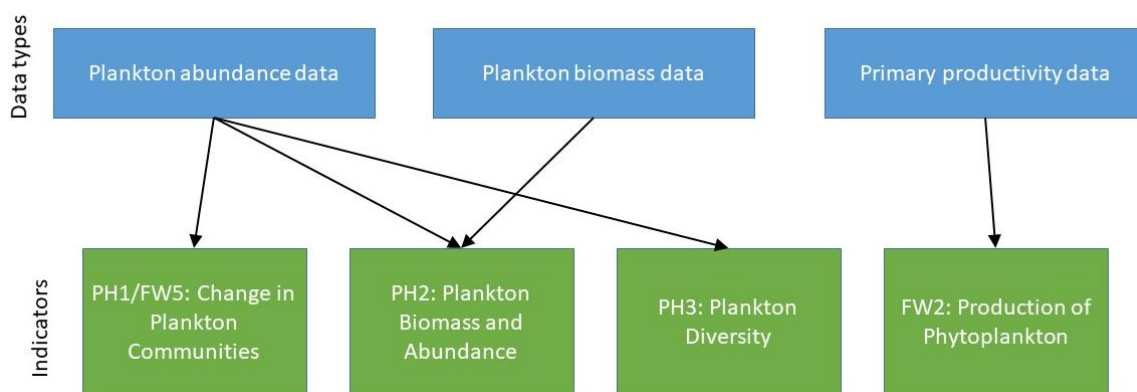


Figure 2. Three types of plankton data can be analysed in multiple ways to inform four indicators. The indicators are flexible and can work with plankton data collected in multiple ways. It is not necessary to supply all data types in response to the data call. Instead, please supply whichever data you would like to submit.

2 Data management responsibilities for data providers

2.1 Plankton abundance data

Phytoplankton and zooplankton abundance data analysed via light microscopy or via image analysis are requested, for PH1/FW5, PH2 (only copepod abundance) and PH3. Submitted data should conform to the highest taxonomic classification possible. Due to the flexible nature of the pelagic indicators, all levels of taxonomic identification are useful, even phylum-, class-, order-, and family-level, so all data is useful, even if not resolved to genus or species.

When submitting taxonomic data, it is important to ensure that no taxa are double-counted and taxonomic identification or enumeration anomalies, and inconsistencies are removed. For example, if a survey records individual *Dinophysis* species as well as a 'Total *Dinophysis*' count, it is necessary to remove the 'Total *Dinophysis*' count if it is simply the sum total of individual *Dinophysis* species. If the 'Total *Dinophysis*' count is calculated another way, a decision will have to be made by the data provider on whether to retain each individual species, the 'Total *Dinophysis*' count, or include another option. For a second example: A survey normally includes all Tintinnids in a 'Total Tintinnid' category. However, for five years a specialist visits the survey and identifies all Tintinnids to species level, but only during that five-year period. In that case, the Tintinnid species categories for the five-year period must be eliminated, as those data are already included in the Total Tintinnid category.

Data used for assessment should not include interpolated species data; if there is a long stretch of time without data for a particular species, this whole species should be removed from the submitted data. Data gaps in the plankton time-series also cause problems for time-series analysis. Plankton abundance follows a seasonal cycle so at least approximately monthly observations are needed to, for example, derive meaningful annual average abundances. The assessment protocol has been designed to accommodate some missing data as well as not to generate outputs when temporal data coverage is inadequate.

Before an institute submits data, Aphia IDs are required to ensure taxonomic consistency, so it is critical that all taxonomic data are linked to Aphia IDs. This can be done through the World Register of Marine Species (WORMS) tool: <http://www.marinespecies.org/aphia.php?p=match> or dedicated R packages 'worms': <https://cran.r-project.org/web/packages/worms/index.html> and 'worrms': <https://cran.r-project.org/web/packages/worrms/index.html>) can partially automate this process.

Individual datasets are then submitted to DASSH, where they are given a unique DOI which can be used to link assessment outputs to specific dataset versions. DASSH holds copies of all datasets underpinning assessment, but these are only made publicly available as raw data if desired by the respective data providers.

Once a dataset is submitted to DASSH, new species will be ingested into the Master Taxa List which assigns biological traits to taxa, allowing them to be sorted into lifeforms for PH1/FW5. The new dataset's species list will be compared with the Master Taxa List via Aphia IDs, and any new species not currently represented on the list will be identified. This process ensures that each species is only entered in the database once. Any new species will be manually assigned functional traits by obtaining expert advice directly from data providers, Pelagic Habitats Expert Groups, and by searching the literature. Data providers may be approached to provide more information if there are taxa in their dataset which are not currently represented on the Master Taxa List. Once traits have been assigned for new species, they can be added to the Master Taxa List. Missing trait information leads to identified taxa not being included in the lifeform analysis.

2.2 Plankton biomass data

PH2 requires phytoplankton biomass and zooplankton abundance (only copepods are considered) data. The requirements for zooplankton abundance data follow the requirements described in the section above, to differentiate copepods from other zooplankton taxa (2.1 *Plankton abundance data*). The indicator for phytoplankton biomass is flexible and can be used to assess multiple data types, including total or size-fractionated chlorophyll *a* concentration by spectrophotometry, extractive fluorometry, *in vivo* fluorometry, remote sensing or by HPLC, phytoplankton biovolume, phytoplankton total and size-class C biomass, etc. If contributing phytoplankton biomass data with a taxonomic component, please follow guidance above regarding providing Aphia IDs and removing

inconsistencies. Individual biomass datasets will also be submitted to and stored on DASSH where they will be tagged with a unique DOI.

Phytoplankton biomass datasets should include interpolated data. Gaps in the plankton time-series also cause problems for temporal analysis. As phytoplankton biomass follows a seasonal cycle, a resolution of at least monthly observations is needed to derive meaningful annually average biomass. The assessment protocol has been designed to accommodate some missing data as well as not to generate outputs when temporal data coverage is inadequate.

When plankton biomass is considered through a taxonomic level, it is necessary to differentiate between phytoplankton and zooplankton. This step requires the Master Taxa List which attributes to each taxa its plankton type (phytoplankton or zooplankton) by matching the Aphia IDs. This task can be processed without using the PLET.

2.3 Primary productivity data

FW2 requires primary productivity data. Phytoplankton primary production can be measured using different methods, sampling strategies and sampling design (see OSPAR, Kromkamp, Capuzzo, & Philippart, 2017). The method adopted to measure production should not affect the assessment as long as the result (Annual Primary Production-APP) is expressed in common units (i.e. $\text{gC}/\text{m}^2/\text{y}$ or $\text{gC}/\text{m}^3/\text{y}$). Therefore, the type of datasets used for FW2 calculation must be consistent. We are looking for data on photosynthetic parameters, addressed by 13C, 14C incubations and/or Oxygen incubations on short time scales but also by variable fluorescence through Fast Light Curves (Fast Repetition Rate fluorescence (FRRf) or Pulse Amplitude Modulated (PAM)): maximum photosynthetic yield (F_v/F_m), relative and/or absolute photosynthetic capacity (P_{max} or ETR_{max}) and photosynthetic efficiency (α), light saturation parameters (E_k) and, when possible, estimates of primary productivity converted into carbon flux from electron flux measurements, at seasonal time scale; NB: the values of electron/carbon conversion factors [K_c or $\phi_{e,c}$] would also be a useful added value for the project. Specific metadata include method, incubation time, light source system and light spectrum for incubation, measuring wavelengths for variable fluorescence etc. Data from primary productivity models and remote sensing are also welcome as they are useful for calculating FW2 at the regional scale. Such as for plankton abundance and biomass, no primary productivity dataset should include interpolated data. Gaps in the time-series also cause problems for temporal analysis. As the primary production follows phytoplankton biomass, a clear seasonal cycle is observed. Thus, a resolution of at least monthly observations is needed to derive meaningful annually average primary production. The assessment protocol has been designed to accommodate some missing data as well as not to generate outputs when temporal data coverage is inadequate.

2.4 Summary of responsibilities for data providers

In summary, the data management responsibilities for data providers are to:

- Ensure species or taxa grouping abundances in their data set are appropriate for aggregation into lifeforms, e.g.: no potential 'double-counting' of individuals,
- Ensure datasets are appropriate for time-series analysis, e.g.: 'changes' in plankton abundance will reflect changes in the plankton community and not changes in monitoring or analysis methodology,
- Compile and provide full metadata to accompany the dataset,
- Provide each species or taxa grouping with an appropriate Aphia ID,
- Format the data into templates provided for submission to DASSH (described below),
- Provide permission for their data to be publicly available from DASSH, if desired,
- Periodically update the dataset so that recent years of observations are available,

- If requested, liaise with Cefas to provide functional trait information for new taxa after submitting data to DASSH.

The OSPAR Pelagic Habitats Expert Group will likely be able to provide data providers with some support for the above tasks.

3 Data call

Data calls will be issued through OSPAR which will serve to 1) update datasets that supported the previous assessment and 2) capture additional datasets that were not yet used in previous assessments (including those using non-microscopy data and satellite data), as well as zooplankton biomass and primary production data (in situ, in situ simulated and also modelling and satellite data).

The process of the data call is straightforward for data supporting the OSPAR pelagic habitats indicators “Change in plankton biomass and abundance” (PH2) and “Change in plankton diversity” (PH3), as well as the food web indicators “Production of phytoplankton” (FW2), “Biomass, species composition and spatial distribution of zooplankton” (FW6) and “Change in plankton communities” (PH1/FW5), which are linked to the food webs FW9 (Ecological Network Analysis) indicator. As an indicator suite, these indicators inform four MSFD Descriptor-Criterion sets: D1C6 (PH1/FW5, PH2, PH3), D4C1 (PH1/FW2 and PH3), D4C2 (PH1/FW5, PH2 and FW6) and D4C4 (FW2).

4 Data reporting format

Along with the data call, guidance will be provided in terms of the data reporting format required for data providers to submit data. Two Excel data template workbooks will be provided to help data providers understand the formatting requirements for data held in matrices and for data held in lists. Data providers should select whichever data template workbook is the best suited for their dataset before submitting to OSPAR. The below tables describe the set of variables required and the correct formatting for plankton abundance data (**Table 1**), plankton biomass data (**Table 2**) and primary productivity data (**Table 3**).

Table 1: Plankton Abundance Matrix Format / Plankton Abundance List Format

Field	Data type	Description	Compulsory /Optional
Contracting Party	Text	The Contracting Party this data is associated with, using ISO two letter country code e.g. DE	Compulsory
Data Access	Text	Select from Public or Restricted. If Restricted, please supply reference material underpinning reasoning (report in “Comment” field) noting The OSPAR Convention Text, Article 9. Access to Information. Default data access will be Public. e.g. Public, Restricted	Optional
Dataset Name	Text	Identifier for your dataset, including your institute name. e.g. PML_L4_Phyto, MBA_CPR Zoo, SAMS_Phyto	Compulsory
Abundance type/units	Text	The type of abundance data you are submitting. Please include units if appropriate.	Compulsory

		e.g. count, cells/ml, SACFOR, presence/absence	
Date	Date	Date of sample e.g. 15/03/2016	Compulsory
Time	Time	Time of sample, hh:mm [24 hrs GMT] e.g. 14:31	Compulsory
Latitude	Number	Latitude of sample, in decimal degrees, calculated using WGS84 e.g. 50.20	Compulsory
Longitude	Number	Longitude of sample, in decimal degrees, calculated using WGS84 e.g. -5.01	Compulsory
Depth (min)	Number	Minimum depth of sample in metres e.g. 7.2	Compulsory
Depth (max)	Number	Maximum depth of sample in metres e.g. 9.3	Compulsory
Size Class	Number	If your phytoplankton taxa has a size component please include in this field. For phytoplankton: 1. >20 um individual cell diameter 2. <20um individual cell diameter	Optional
Taxon	Text	Name of taxon e.g. <i>Crepidula grandis</i> , Copepod, Ceratium spp., Centric diatom	Compulsory
aphiaID	Number	Aphia ID for taxon at the appropriate level. It is not necessary for all taxa submitted to be of the same taxonomic resolution, but the higher the resolution, the better. Aphia IDs can be found at http://www.marinespecies.org/aphia.php?p=match e.g. 254469, 1080, 109506, 149013	Compulsory
Abundance	Number	Measured plankton parameter, corresponding to the abundance type and units in the Abundance type/units field. e.g. 5 (for count data), Present (for presence/absence data), C (for SACFOR data)	Compulsory
Comment	Text	Additional information linked to the reporting. Note, if the dataset is marked as restricted, the reasoning needs to be defined here	Optional

Table 2: Plankton Biomass Matrix Format / Plankton Biomass List Format

Field	Data type	Description	Compulsory /Optional
Contracting Party	Text	The Contracting Party this data is associated with, using ISO two letter country code e.g. DE	Compulsory

Data Access	Text	Select from Public or Restricted. If Restricted, please supply reference material underpinning reasoning (report in "Comment" field) noting The OSPAR Convention Text, Article 9. Access to Information. Default data access will be Public. e.g. Public, Restricted	Optional
Dataset Name	Text	Identifier for your dataset, preferably including your institute name. e.g. PML_L4_Phyto, MBA_CPR Zoo, SAMS_Phyto	Compulsory
Plankton Biomass Parameter Type and Units	Text	The type of biomass data you are submitting. Please include units if appropriate. e.g. chlorophyll <i>a</i> (spectrophotometer, fluorimeter, HPLC, $\mu\text{g m}^{-3}$), Phytoplankton Colour Index, Phyto- or Zooplankton biovolume ($\text{mm}^3 \text{m}^{-3}$), Phyto- or Zooplankton biomass (mg m^{-3})	Compulsory
Date	Date	Date of sample e.g. 15/03/2016	Compulsory
Time	Time	Time of sample, hh:mm [24 hrs GMT] e.g. 14:31	Compulsory
Latitude	Number	Latitude of sample, in decimal degrees, calculated using WGS84 e.g. 50.20	Compulsory
Longitude	Number	Longitude of sample, in decimal degrees, calculated using WGS84 e.g. -5.01	Compulsory
Depth (min)	Number	Minimum depth of sample in metres e.g. 7.2	Compulsory
Depth (max)	Number	Maximum depth of sample in metres e.g. 9.3	Compulsory
Size Class	Number	If your phytoplankton taxa has a size component please include in this field. For phytoplankton: 1. >20 um individual cell diameter 2. <20um individual cell diameter	Optional
Biomass parameter	Text	Name of biomass parameter e.g. <i>Crepidula grandis</i> , Copepod, Ceratium spp., Centric diatom, Chlorophyll <i>a</i>	Compulsory
aphiaID	Number	Aphia ID for taxon at the appropriate level. It is not necessary for all taxa submitted to be of the same taxonomic resolution, but the higher the resolution, the better. Aphia IDs can be found at http://www.marinespecies.org/aphia.php?p=match e.g. 254469, 1080, 109506, 149013	Compulsory if submitting taxonomic biomass data
Biomass	Number	Measured plankton biomass parameter, corresponding to the biomass data type and units in the Biomass type/units field	Compulsory

		e.g. 5, 231, 0.1	
Comment	Text	Additional information linked to the reporting. Note, if the dataset is marked as restricted, the reasoning needs to be defined here	Optional

Table 3: Primary Productivity Format

Field	Data type	Description	Compulsory /Optional
Contracting Party	Text	The Contracting Party this data is associated with, using ISO two letter country code e.g. DE	Compulsory
Data Access	Text	Select from Public or Restricted. If Restricted, please supply reference material underpinning reasoning (report in "Comment" field) noting The OSPAR Convention Text, Article 9. Access to Information. Default data access will be Public. e.g. Public, Restricted	Optional
Dataset Name	Text	Identifier for your dataset, preferably including your institute name. e.g. PML_L4_Phyto, MBA_CPR Zoo, SAMS_Phyto	Compulsory
Phytoplankton Primary productivity and photosynthetic parameters Type/Units	Text	The type of phytoplankton primary productivity data you are submitting (measured PP or photosynthetic parameters and/or computed PP from photosynthetic parameters or models at ecosystem level). Please include measured variable (and for VF the ETR algorithm computation used), incubation technique (time and light source), data correction method, FLC fit model, FLC quality parameter, units if appropriate. e.g. In situ Primary production measurements (Carbon absorption, oxygen flux); or photosynthetic parameters (Fv/Fm, r.alpha, alphaII, P_{max}^B , $r.ETR_{max}$, $ETR_{II_{max}}$, r.Ek, etc. see below) from C incubations or from variable fluorescence (VF, Fast Repetition Rate fluorescence, PAM, PhytoPAM tech.), 15 mn under artificial white or blue light , VF background correction (with filtered sea water on GF/F), Eilers and Peters (1988) model, R ² fit control; or PP from ecosystem models	Compulsory
Date	Date	Date of sample e.g. 15/03/2016	Compulsory
Time	Time	Time of sample, hh:mm [24 hrs GMT] e.g. 14:31	Compulsory
Latitude	Number	Latitude of sample, in decimal degrees, calculated using WGS84 e.g. 50.20	Compulsory

Longitude	Number	Longitude of sample, in decimal degrees, calculated using WGS84 e.g. -5.01	Compulsory
Depth (min)	Number	Minimum depth of sample in metres e.g. 7.2	Compulsory
Depth (max)	Number	Maximum depth of sample in metres e.g. 9.3	Compulsory
Primary Productivity	Number	Surface or vertically integrated PP rates on daily, monthly or annual time scale (in units and from technique previously described) e.g. 600 $\mu\text{g C}\cdot\text{m}^{-2}\cdot\text{h}^{-1}$	Compulsory
Maximum Yield F_v/F_m	Number	Quantum efficiency of PSII under dark conditions e.g. $F_v/F_m = 0.56$ (dimensionless, 10' dark acclimation)	Optional
r.alpha, alpha _{all} or alpha _{RCII}	Number	Initial slope of photosynthesis light curve (from C, O ₂ or VF technique) in relative or absolute units (specify the absorption variable studied and units) e.g. $\alpha_{\text{RCII}} = 2.5 \text{ mol electrons}\cdot\text{mol RCII}^{-1}\cdot\text{s}^{-1}$ (using RCII number, $a^*\text{Phy}$, a_{LHII} absorption or SigmaPSII)	Optional
$P^{\text{B}}_{\text{max}}$ or r.ETR _{max} , ETR _{max} or ETR _{RCIImax}	Number	Photosynthetic capacity/carbon assimilation number (C or O ₂ technique) or maximum photosynthetic electron transport rate in relative or absolute units (VF with the same units and technique as alpha _{all}) e.g. $P^{\text{B}}_{\text{max}} = 5.5 \text{ mgC}\cdot(\text{mgChla})^{-1}\cdot\text{h}^{-1}$ or $\text{ETR}_{\text{RCIImax}} = 350 \text{ mol electrons}\cdot\text{mol RCII}^{-1}\cdot\text{s}^{-1}$	Optional
r.Ek, Ek _{II} , or Ek _{RCII}	Number	Saturation irradiance parameter in relative or absolute units as previous parameters e.g. $\text{Ek}_{\text{RCII}} = 270 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$	Optional
Other parameters	Number	Like $\Phi_{e,c}$ (electron/carbon conversion factors) and photobiological parameters: NPQ (Non Photochemical Quenching), a^*_{phy} (specific absorption coefficient of phytoplankton), a_{LHII} (absorption coefficient of PSII light harvesting), SigmaPSII (functional absorption cross section of PSII), RCII concentration (Reaction Center II concentration or number) etc e.g. $\Phi_{e,c} = 18.9 \text{ mole}^{-1}\cdot\text{mol C}^{-1}$	Optional

5 References

- Kromkamp, J., Capuzzo, E., & Philippart, C. (2017). Measuring phytoplankton primary production: review of existing methodologies and suggestions for a common approach. EcApRHA Deliverable WP 3.2.
- Ostle, C., Paxman, K., Graves, C. A., Arnold, M., Artigas, L. F., Atkinson, A., . . . Bedford, J. (2021). The Plankton Lifeform Extraction Tool: a digital tool to increase the discoverability and usability of plankton time-series data. *Earth System Science Data*, 13(12), 5617-5642.