Framework Service Contract EEA/DIS/R0/20/001 Lot 1 for Services supporting the European Environment Agency's (EEA) cross-cutting coordination of the Copernicus programme's in situ data activities – Observational data

Copernicus requirements for Biogeochemistry Essential Ocean Variables in the coastal ocean

Lőrinc Mészáros
Jun She
Ghada El Serafy

Issue:
Date: 28-10-2021

[Document Release]

	Name(s)	Affiliation
Coordinated by:	Lorinc Meszaros, Ghada El Serafy	Deltares
Contributions:	Jun She	Danish Meteorological Institute
Approval:		

Prepared for:	European Environment Agency (EEA)
Represented by: (Project Manager)	Henrik Steen Andersen
Contract No.	EEA/DIS/R0/20/001 Lot 1

[Change Record]

Version	Date	Changes
1.0	28-2-2021	First draft release
1.1	10-3-2021	First revision
1.2	7-4-2021	Second revision based on comments from the EuroGOOS Office
1.3	13-4-2021	Final revision
1.4	8-10-2021	Future requirements of the IBI and BS MFCs added and the current status of the ARC MFC updated. Other updates in the executive summary, and recommendations sections.
1.5	28-10-2021	Final revision based on comments from the EuroGOOS Office

Table of Contents

Executive Summary

This report contributes to the EEA framework contract 'Support to Copernicus In Situ Data Coordination' (COINS) and specifically targets Task 1.1:' Identify and create a list of user requirements (resolution in time and space, quality, timeliness) for Biogeochemistry (BGC) Essential Ocean Variables (EOVs) data in the coastal zone for Copernicus'.

Within this task, the CMEMS Monitoring Forecasting Centers (MFCs) were approached with a survey, approved and facilitated by Mercator Ocean International, to report on their requirements regarding BGC EOVs. The survey covered two aspects of the future requirements: purpose of use and data characteristics. The Marine Forecasting Systems were asked about their in-situ data requirements for algorithm development, model calibration and validation, and data assimilation. Moreover, they were questioned about the characteristics of the required in-situ data, such as the number of stations, the area covered, update frequency, timeliness, and accuracy. The MFCs cover regional seas and serve as an intermediate step towards higher resolution coastal models (as boundary conditions and initial conditions), therefore the regionals MFCs have been contacted to express their requirements to improve coastal modelling capacities. The compiled feedbacks from the MFCs and the synthesis of their commonalities and gaps aims to help the EEA to better formulate long term in-situ observations coordination strategies, which is in the interest of CMEMS and its MFCs and Thematic Assembly Centers (TACs).

Results from the survey indicate that turbidity and phytoplankton biomass are among the BGC in-situ variables requested by MFCs, which are currently not used or only used to a limited extent. This may indicate that MFCs aim to further develop modelling systems and improve prediction skills for turbidity and phytoplankton biomass simulations. Thus, providing in-situ observations of these variables may be an enabling factor towards improved BGC modelling capacity in the coastal regions. Regarding the spatial resolution and coverage, the expressed requirements are largely heterogenous varying from 5 to 100 km of spatial resolution requirements. Some MFCs estimated that they need 50 locations in their basin for model validation, while others didn't specify the spatial density but requested that monitoring stations should cover all parts of the basin (to ensure a good spatial coverage) or specifically target major river mouths. It was found that the lack of BGC in-situ data is problematic for the MFCs and the evaluation of the quality of the model products is hindered by the poor observational coverage.

In addition, the current use of BGC in situ observations are not focused on the coastal areas but target the whole basin. This is partially due to the existing BGC data sources, such as BGC Argo floats, which cover the open ocean rather than coastal areas. Moreover, in the coastal regions, the MFCs only play an intermediate role towards the higher resolution coastal models by providing boundary conditions and initial conditions. Improving and validating the forecasting skills (product quality) of CMEMS MFCs in coastal areas to better serve the coastal modelling community will therefore require concentrated effort on the coast. Consequently, one of the recommendations is that more operational BGC in-situ data sources are needed in the coastal area and especially near major river mouths. Apart from operational in-situ fixed platforms, such as moorings or surface Ferry Boxes in ships of opportunity, periodic missions that also provide profiles (e.g., gliders) are also crucial to allow vertical validation of the biogeochemical models.

Concerning the data quality and accuracy requirements, it was expressed by the MFCs that in general bias and uncertainty estimates are very helpful but, in most cases, there were no in-situ data accuracy requirements formulated, except in the Arctic MFC where the expected accuracy of the chlorophyll-a observations is 10%. In other words, information on measurement precision (measurement

uncertainty) is more important than precise information on its accuracy. With regards to the temporal requirements, most BGC in-situ observations are required weekly for product monitoring or with a lower update frequency (monthly or six-monthly) to serve interim production and reanalysis purposes. Finally, improving the near real time delivery for chlorophyll-a, dissolved oxygen, pCO2, particulate matter, and pH should be a priority in coastal areas.

Introduction

In 2018, a report[1] was jointly prepared by Mercator Ocean, EUROGOOS, and CMEMS partners. The report was entitled 'CMEMS requirements for the evolution of the Copernicus In Situ Component'. In this report the main CMEMS requirements for the evolution of the Copernicus in-situ component have been summarized after a dedicated workshop held in Toulouse in July 2018. The report concluded the following requirements (related to BGC variables):

- There are critical sustainability gaps in the in-situ observation network and major gaps for biogeochemical observations (carbon, oxygen, nutrients, Chl-a). New mechanisms need to be set up between the EU and member states to address them.
- Following the evolution of CMEMS models in term of spatial resolution, which in future will reach the kilometric scale at global level, there is a clear need of more sensors deployed at global and regional scale.
- Improved timeliness in data delivery is required to ensure that data are available at each model run; particularly for coastal applications.
- Consolidation of the Argo mission including the sampling of polar seas and marginal seas and developing the BGC Argo is a strong priority.
- Improving key observing systems such as ferry-boxes, gliders, are strong priorities for regional CMEMS products.
- A specific effort for the Arctic region is needed; there are severe limitations with measurements over the seasonal ice zone. More ITPs, core and BGC Argo floats are needed.
- Data harmonization and access need to be improved using distributed and connected databases.

In this report we don't intend to reproduce the same discussion but rather to complement it with a more holistic view on the biogeochemical aspects on the regional seas, which serve as crucial intermediate step to be able to target the coast.

The focus of this report is therefore on the regional level (European Regional Seas) represented by the CMEMS MFCs instead of the coastal level. Describing BGC requirements for the regional level is essential to enable better biogeochemical coastal forecasting, as CMEMS MFCs provide the boundaries and initial conditions for higher resolution coastal models. Consequently, regional MFCs have been contacted to express their requirements to improve coastal modelling. We assume that Copernicus requirements for Biogeochemical Essential Ocean Variables in the coastal ocean can be mainly described by the individual requirements of the regional CMEMS Monitoring Forecasting Centres (MFCs) for data assimilation and model calibration/validation.

[1] https://marine.copernicus.eu/sites/default/files/media/pdf/2020-10/CMEMS-requirements-In_Situ.pdf

Existing and future requirements of Biogeochemical EOVs per CMEMS Monitoring and Forecasting Centre (MFCs)

In this section we first investigate the simulated (forecast or analysis) BGC variables and the current use of BGC in-situ observations BGC EOVs by the regional MFCs in their forecast and hindcast products by extracting the information provided in the Quality Information Documents (QUIDs) accessible on the CMEMS Website (http://marine.copernicus.eu).

The current use of BGC EOVs in situ observations is then complemented with an analysis on the future requirements based on results from a survey conducted within this task. CMEMS MFCs were approached with a survey, approved and facilitated by Mercator Ocean International, where they were asked to report on their future requirements regarding BGC EOVs in situ observations. The survey covered two aspects: which is the purpose of use and data characteristics (variables, resolution, updated frequency, accuracy etc.). The survey template is attached to this report as an Appendix.

Global MFC

Future requirements of the Global Monitoring and Forecasting Center (GLO MFC) are connected to the BGC forecast and hindcast products. In the hindcast product currently in-situ BGC observations are only used for validation, whereas in the forecast product, only satellite observed chlorophyll is used for data assimilation. Therefore, in the Global MFC BGC in-situ data is currently not used for data assimilation.
Both forecast and hindcast products provide data for following variables:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_phytoplankton_expressed_as_carbon_in_sea_water (PHYC)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- mole_concentration_of_silicate_in_sea_water (SI)
- mole_concentration_of_dissolved_iron_in_sea_water (FE)
- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)
- sea_water_ph_reported_on_total_scale (PH)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_water (PP)

Validated BGC variables are chlorophyll, nitrate, phosphate, silicate, dissolved oxygen, phytoplankton carbon concentration, primary production, and pH (for the forecast product). As an example, *Figure 1* shows the simulated surface chlorophyll-a concentration against BGC-Argo floats.

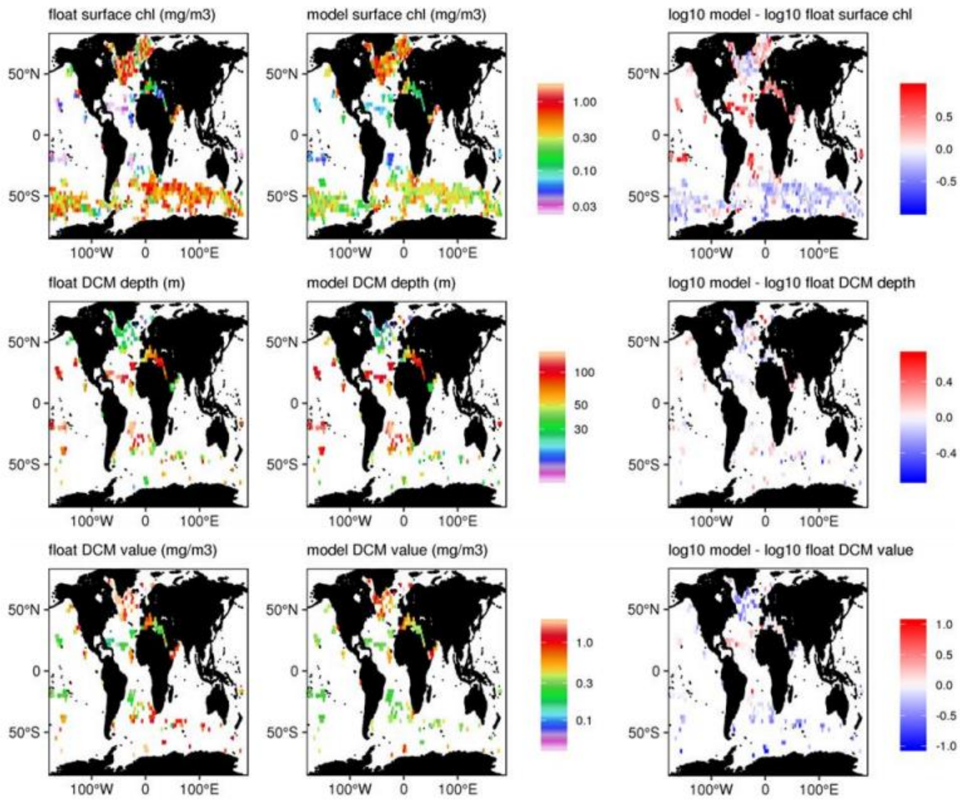

Figure 1. Surface chlorophyll validation against BGC-Argo floats between 2008-2017. (Source : https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-GLO-QUID-001-028.pdf)

Future requirements

Future requirements of the Global MFC have not been evaluated in the framework of this report as the task focuses on European Regional Seas.

Arctic MFC

Future requirements of the [Arctic Ocean MFC (ARCMFC)]() are connected to the BGC forecast and reanalysis products. In the forecast product in-situ BGC observations are only used for validation, whereas satellite observed chlorophyll is used for data assimilation. In the reanalysis product, nutrients data is assimilated (nitrate, phosphate, silicate) from in-situ observations and surface chlorophyll is assimilated from satellite sources. Therefore, in the Arctic MFC BGC in-situ data is used for both validation and data assimilation, thanks to recent upgrades in their modelling system.

Forecast and hindcast products provide BGC data for the following parameters:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_phytoplankton_expressed_as_carbon_in_sea_water (PHYC)
- mole_concentration_of_zooplankton_expressed_as_carbon_in_sea_water (ZOOC)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- mole_concentration_of_silicate_in_sea_water (SI)
- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)

- sea_water_ph_reported_on_total_scale (PH)
- mole_concentration_of_dissolved_inorganic_carbon_in_sea_water (DIC)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_wate r (PP)
- volume_attenuation_coefficient_of_downwelling_radiative_flux_in_sea_water (KD)

In the forecast product only nitrate, phosphate, and silicate are validated against in-situ data as primary production, phytoplankton, zooplankton, oxygen, attenuation and pH have insufficient data. As mentioned above, in-situ BGC observations are only used for validation in the forecast product. Chlorophyll-a is validated against satellite products. In-situ observations for nitrate, silicates and phosphate are not available for the recent years and for this reason the year 2015 was used as a reference year for the nutrient validations. Moreover, it should be noted that in the Arctic MFC even the chlorophyll-a satellite product has high uncertainty due to the local conditions (cloudiness and sun position). **Therefore, BGC in-situ observations in this region are crucial and, in many cases, irreplaceable.**

The sources of in-situ nutrient observations are the Institute of Marine Research (IMR, Norway), World Ocean Atlas 2013 dataset for the forecast product, GLODAPv2_2019, ICES oceanographic database and CLIVAR datasets for the reanalysis product. Even after combining these data sources the number of in-situ nutrient data are limited in the model domain. The number of observation data assimilated at each reanalysis cycle in 2007 is depicted in the figure below.

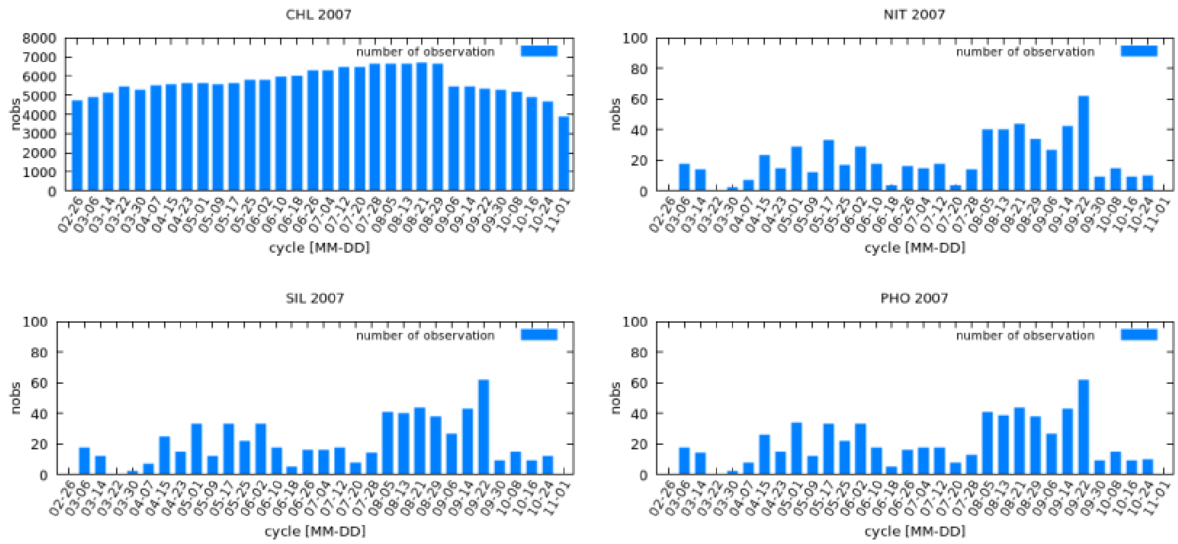

Figure 2. Number of observation data assimilated at each reanalysis cycle in 2007. Surface chlorophyll-a from satellite and nutrients from in-situ observations (Source: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-ARC-QUID-002-005.pdf)

Future requirements

The Arctic MFC reported BGC in-situ data requirements on chlorophyll-a, dissolved oxygen, pCO2, pH, dissolved N, dissolved P, dissolved Si, primary production for both model calibration/validation and data assimilation. Moreover, they require turbidity, phytoplankton diversity and biomass, and zooplankton diversity and biomass for model calibration/validation.

In terms of spatial coverage, profiles are required every 100 km for the entire Arctic, and at least 50 stations are needed. Chlorophyll-a is required at a higher resolution at the surface. Currently there exist big data gaps in the non-European Arctic. The QUID[2] document states that in general the observational coverage (space and time) is inadequate to provide a comprehensive evaluation of the forecast product quality. The required update frequency for all variables is weekly, with 3 to 6 months delay (timeliness). The expected accuracy of the chlorophyll-a observations is 10%, this requirement may be explained by the associated high uncertainty of the chlorophyll-a satellite product due to the local conditions (cloudiness and sun position).

Baltic MFC

Future requirements of the Baltic Sea MFC (BAL MFC) are connected to the BGC forecast and hindcast products. In the forecast product in-situ BGC observations are only used for validation, whereas in the hindcast product, oxygen, and nutrients (nitrate, phosphate, ammonium) are being used for data assimilation.
Both forecast and hindcast products provide data for following parameters:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- mole_concentration_of_ammonium_in_sea_water (NH4)

In the forecast product the following additional BGC variables are included:

- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)
- sea_water_ph_reported_on_total_scale (PH)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_water (PP)
- secchi_depth_of_sea_water (ZSD)

In-situ BGC variables being used for validation on both forecast and hindcast products are dissolved oxygen, ammonium, nitrate, phosphate, chlorophyll-a. Additional in-situ BGC variables in the forecast product are Secchi depth, pH, pCO2. In-situ BGC data for the BAL MFC is depicted in *Figure 3*.

2 https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-ARC-QUID-002-004.pdf

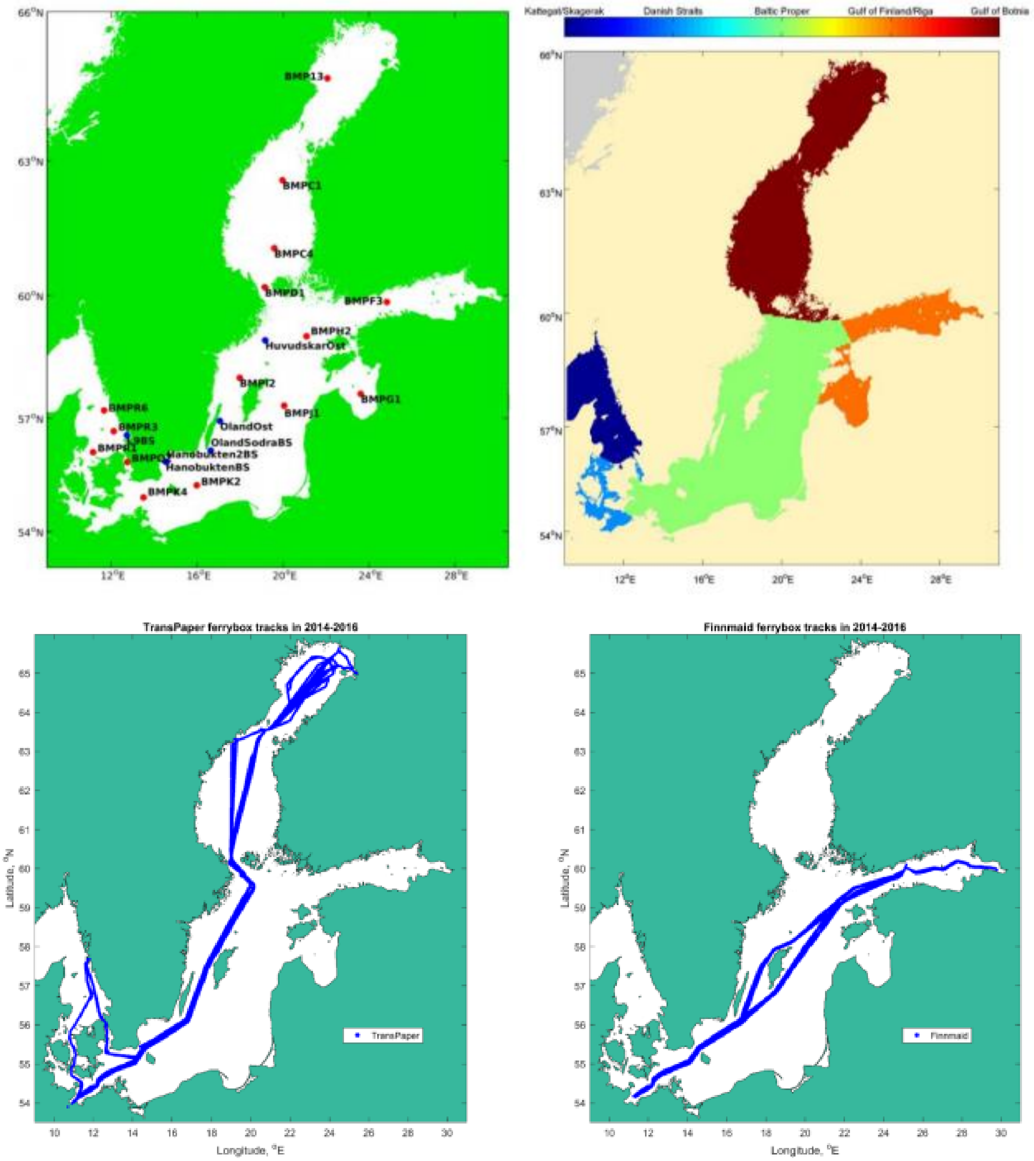

Figure 3. In-situ BGC data being used for validation in the BAL MFC. Upper panel: All moorings (blue dots) and monitoring stations (red dots) that are accounted for the product assessment (left) all five sub regions, which are used for location of the monitoring stations and moorings (right) Lower panel: Locations of Ferry box observations along the tracks of the ships for the hind-cast period 1st Oct 2014 to 30th Sep 2016. Left: Locations of dissolved oxygen measurements; right: Locations of pCO2 measurements. (Source: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-BAL-QUID-003-007.pdf)

Future requirements

The Baltic MFC requires BGC in-situ data on chlorophyll-a, dissolved oxygen, and turbidity for model calibration/validation and data assimilation. Additionally, Secchi depth, dissolved nutrients, ammonium, pCO2, and pH is required for model calibration/validation. Phytoplankton diversity and biomass in-situ data is needed for algorithm development.

The required spatial coverage for this MFC is the full region of the North and Baltic Seas. The required update frequency for most variables is daily (Secchi depth, dissolved nutrients, ammonium, pCO2, pH, Phytoplankton diversity and biomass), whereas turbidity, chlorophyll-a, and dissolved oxygen is required 6-hourly. In terms of timeliness, there is no special requirement for Secchi depth, pCO2, pH, and Phytoplankton diversity and biomass. On the other hand, turbidity, chlorophyll-a, dissolved oxygen, dissolved nutrients, and ammonium are required in near-real time to allow operational data assimilation. Following CMEMS terminology, near-real time means less than one day after acquisition.

Currently, coastal zone BGC data have not been widely used for model validation. Considering the different characteristics of chlorophyll-a and nutrients in the coastal and open waters (e.g., stratification, vertical gradients, or peak concentrations), BGC observations in the coastal zone will be important for assessing model product quality in the coastal waters.

North Western Shelf MFC

Future BGC requirements of the Atlantic European NorthWest shelves (NWS MFC) are connected to the BGC forecast and hindcast products. In both the forecast and hindcast products BGC observations (satellite measured chlorophyll) are used for data assimilation.

Both forecast and hindcast products provide data for following parameters:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_phytoplankton_expressed_as_carbon_in_sea_water (PHYC)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)
- sea_water_ph_reported_on_total_scale (PH)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_water (PP)
- volume_beam_attenuation_coefficient_of_radiative_flux_in_sea_water (KD)

In the reanalysis product the following additional BGC variables are included:

- mass_concentration_of_picophytoplankton_expressed_as_chlorophyll_in_sea_water (PSC)
- mass_concentration_of_diatoms_expressed_as_chlorophyll_in_sea_water (PFT)

The following BGC variables are being validated with in-situ observations in both forecast and hindcast products: chlorophyll, nitrate, phosphate, oxygen, pCO2, and pH. An additional BGC variable being validated against in-situ data in the forecast product is light attenuation. *Figure 4* shows an example of in-situ BGC observations used for model validation.

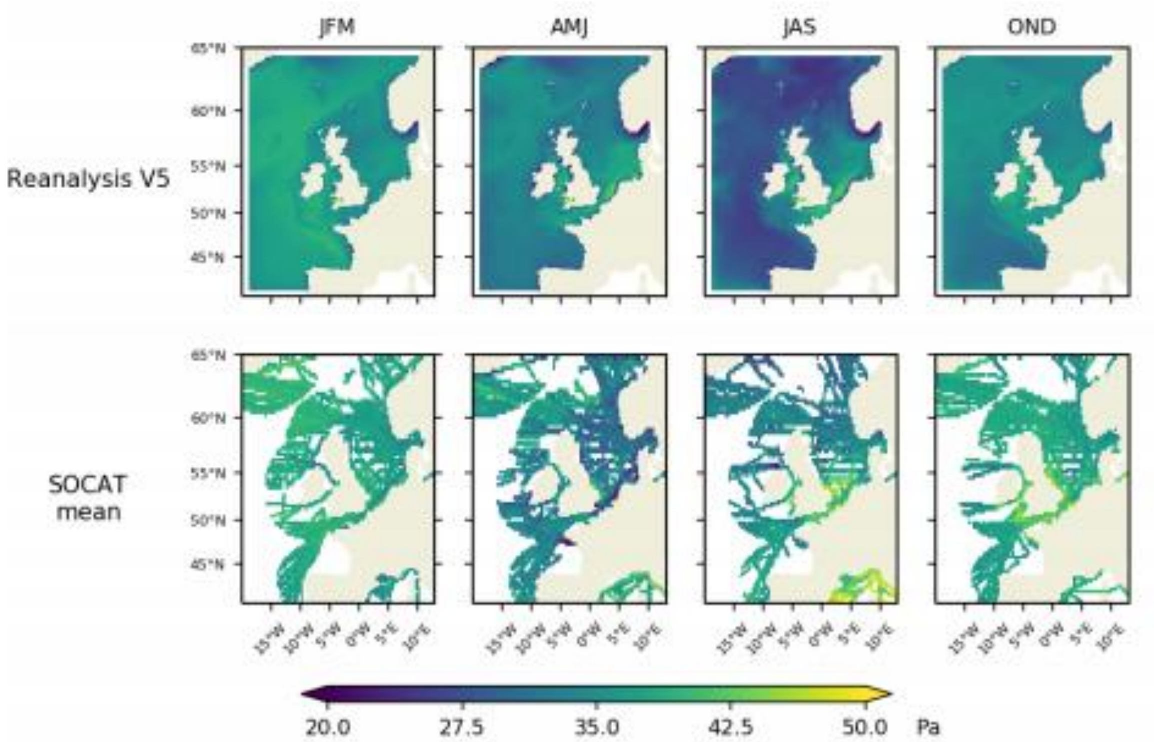

Figure 4. Seasonal average surface pCO2 from the model (top row) and in situ observations from the SOCAT database (bottom row). (Source: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-NWS-QUID-004-011.pdf)

Future requirements

For the NWS MFC, required BGC in-situ data for algorithm development is turbidity, chlorophyll-a, dissolved oxygen, dissolved nutrients, and Phytoplankton diversity and biomass with higher priority; as well as ammonium, particulate matter, Dissolved organic carbon (DOC), Particulate Organic Carbon (POC), and Zooplankton diversity and biomass with lower priority.

For calibration/validation Secchi depth, ammonium, particulate matter, dissolved organic carbon (DOC), and Particulate Organic Carbon (POC) are required with less priority; Zooplankton diversity and biomass is required with medium priority; and turbidity, chlorophyll-a, dissolved oxygen, dissolved nutrients, pCO2, pH, and Phytoplankton diversity and biomass are required with the highest priority and have the biggest impact on quality.

For data assimilation, ammonium, particulate matter, Dissolved Organic Carbon (DOC), and Particulate Organic Carbon (POC), and Zooplankton diversity and biomass are required with lower priority; turbidity and Phytoplankton diversity and biomass are required with medium priority; and chlorophyll-a, dissolved oxygen, dissolved nutrients, pCO2, and pH are required with the highest priority and have the biggest impact on quality.

The ideal spatial coverage requirement includes data from all parts of the NorthWest Shelf region, including areas not immediately next to a coastline. Any data update frequency is useful but having data in all seasons is a requirement. The data timeliness is ideally by 5am next day for assimilation, or along the following day; within a few days for product monitoring; and within a few months for model validation. An estimate of bias and measurement uncertainty is deemed to be very helpful for the NWS MFC, while no absolute in-situ data accuracy requirements were formulated.

Iberian-Biscay-Irish Seas MFC

The following evaluation of using BGC observations in CMEMS Iberian Biscay Irish Seas (IBI MFC) are made based on the QUID reports of corresponding BGC forecast and hindcast products.
In situ BGC observations are used in IBI MFC only for validation. No BGC in situ observations are being assimilated.

Both forecast and hindcast products provide data for following parameters:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_phytoplankton_expressed_as_carbon_in_sea_water (PHYC)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- mole_concentration_of_silicate_in_sea_water (SI)
- mole_concentration_of_dissolved_iron_in_sea_water (FE)
- mole_concentration_of_ammonium_in_sea_water (NH4)
- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)
- sea_water_ph_reported_on_total_scale (PH)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_water (PP)
- euphotic_zone_depth (ZEU)
- mole_concentration_of_dissolved_inorganic_carbon_in_sea_water (DIC)

Among the above model parameters, Chlorophyll, Primary productivity, Euphotic layer depth, Nitrates, Phosphate, Silicate, Oxygen, spCO2, pH and Dissolved Inorganic Carbon are validated against in-situ observations. Phytoplankton, iron and ammonium are not validated against in-situ data.
The in-situ data used for validation (coming from the GLODAP database - https://www.glodap.info/) is shown in *Figure 5*. All stations have long-term observations for the period between 1993-2016.

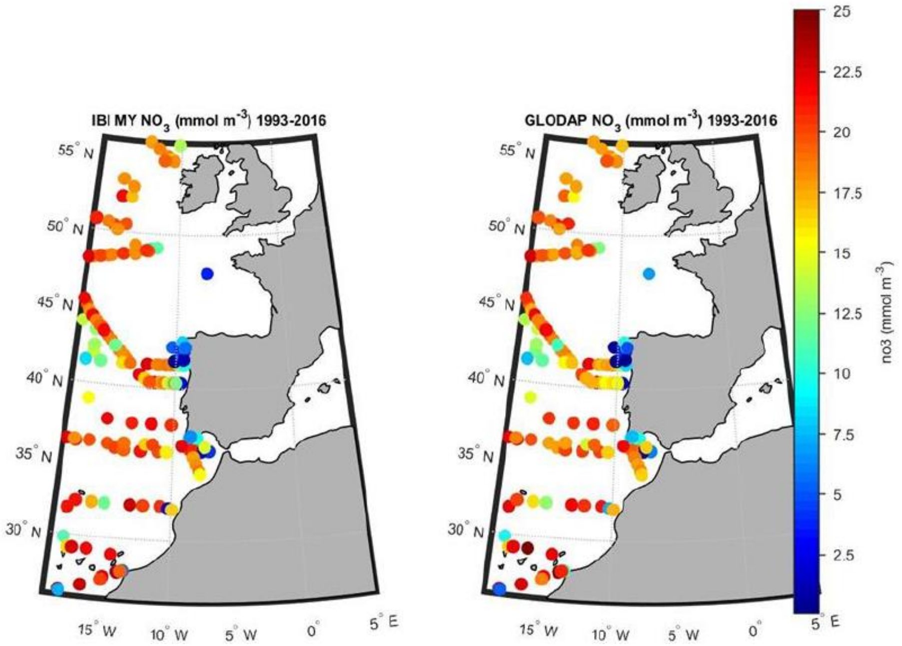

Figure 5. Depth-averaged NO_3 predicted by IBI Biochemical Multi-Year Product (left) against GLODAP NO_3 in-situ (right). (Source: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-IBI-QUID-005-003.pdf)

Future requirements

The IBI MFC requires chlorophyll-a, dissolved nutrient, pCO2 and phytoplankton data for data assimilation purposes, whereas turbidity, secchi depth and chlorophyll-a variables are also required for algorithm development. For model calibration and validation, the IBI MFC requires all listed BGC variables except turbidity and secchi depth. For the development of Ocean Monitoring Indicators, the following BGC variables are required: turbidity, secchi depth, chlorophyll-a, dissolved oxygen, dissolved nutrient, pCO2, pH, phytoplankton, and zooplankton.

Regarding future spatial coverage, all BGC observations should focus on the continental shelf, especially at mouths of major rivers. According to the existing validation of the reanalysis product, major geographical gaps of in-situ observations are found in Bay of Biscay (*Figure 5*). In terms of temporal requirements, at least weekly frequency is requested for all BGC variables. These should be available throughout the year for dissolved oxygen, dissolved nutrients, ammonium, pCO2, Particulate matter, DOC, POC and pH. On the other hand, for optical and biological variables the emphasis is on spring and autumn: turbidity, secchi depth, chlorophyll-a, phytoplankton, and zooplankton. The IBI MFC requests the best possible observational accuracy.

Mediterranean MFC

Future BGC requirements of the Mediterranean Sea MFC (MED MFC) are connected to the BGC forecast and hindcast products. In both, forecast and hindcast products, BGC in situ observations are used for data assimilation. These are chlorophyll and phytoplankton groups for both hindcast and forecast products and oxygen, nutrient concentrations (nitrate, phosphate) for the forecast product.

Both forecast and hindcast products provide data for following parameters:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_phytoplankton_expressed_as_carbon_in_sea_water (PHYC)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)
- sea_water_ph_reported_on_total_scale (PH)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_water (PP)

In the forecast product the following additional BGC variable is included:

- surface_downward_mass_flux_of_carbon_dioxide_expressed_as_carbon (FGCO2)

In the reanalysis product the following additional BGC variable is included:

- mole_concentration_of_dissolved_inorganic_carbon_in_sea_water ()

BGC variables validated against in-situ observations in both forecast and hindcast products are chlorophyll, net primary production, phosphate, nitrate, oxygen, pH, pCO2, DIC, and Alkalinity. Additional BGC variable in the forecast product validated against in situ observations is CO2 air-sea flux. *Figure 6* shows BGC in-situ data sources for validation, such as the NODC-OGS data and the costal monitoring network from ISPRA.

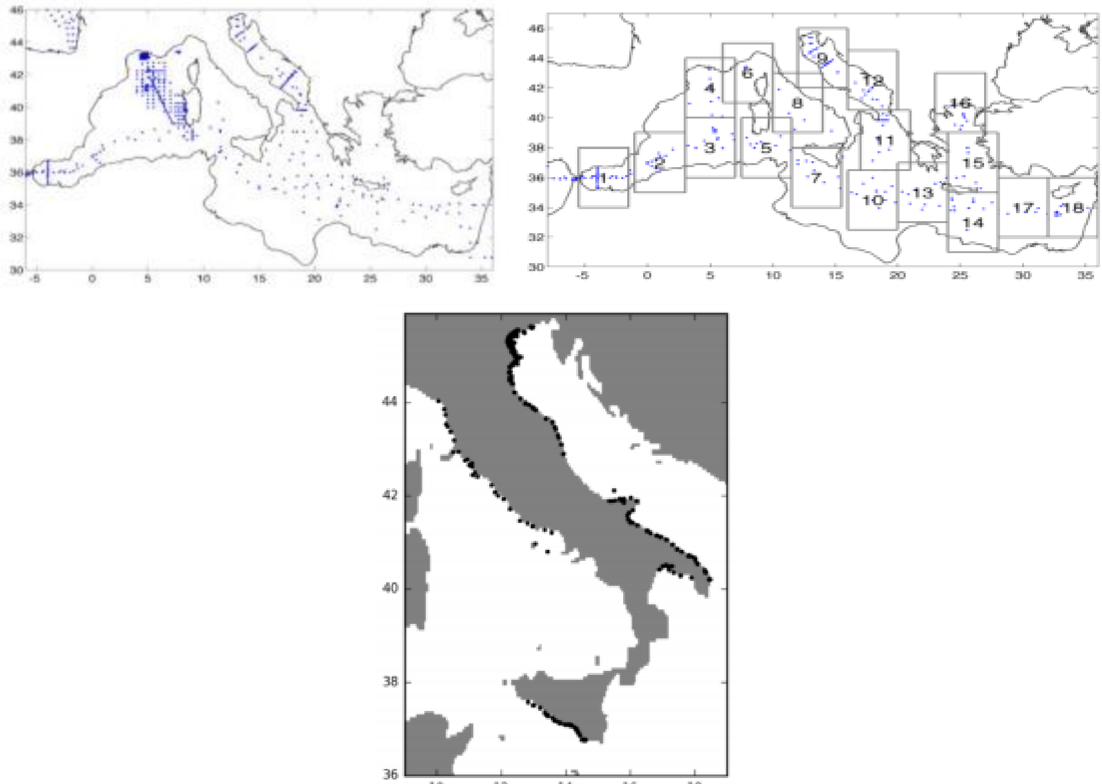

Figure 6. BGC in-situ validation data in the MED MFC. Top left: Map of gathered data of nutrients through the NODC-OGS data set for period 2000- 2011. Top right: Location of the carbonate system variables data. Bottom: Location of the coastal monitoring network of ISPRA (Source: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-MED-QUID-006-008.pdf)

Future requirements

The MED MFC requires BGC in-situ data on chlorophyll-a, dissolved oxygen, dissolved nutrients for model calibration/validation and data assimilation. Moreover, they require ammonium, pCO2, Particulate matter, Dissolved organic carbon (DOC), Particulate Organic Carbon (POC), pH, phytoplankton diversity and biomass, and zooplankton diversity and biomass only for model calibration/validation.

In terms of spatial coverage, stations are required every 10 km of the coast for the whole Mediterranean coastline or, at least, on coastal areas near rivers (e.g., Adriatic Sea, Aegean Sea, France and Spain coastlines). The required update frequency for chlorophyll-a, dissolved oxygen, dissolved nutrients, ammonium, pCO2, Particulate matter, Dissolved organic carbon (DOC), Particulate Organic Carbon (POC), pH is monthly (or every 6 months) to serve interim production or reanalysis. Near real time data (less than one day after acquisition) from automatic sensors is required for chlorophyll-a, dissolved oxygen, pCO2, Particulate matter, and pH. For the rest of requested variables delayed mode is sufficient.

Black Sea MFC

The following evaluation of the use of BGC in situ observations in CMEMS Black Sea MFC (BS MFC) are made based on the QUID reports of corresponding BGC forecast and hindcast products.

In the forecast products, satellite measured chlorophyll is used for data assimilation. No BGC in situ data have been assimilated in the reanalysis product. The output variables of the MFC forecast and hindcast products are listed below.

Both forecast and hindcast products provide data for following parameters:

- mass_concentration_of_chlorophyll_a_in_sea_water (CHL)
- mole_concentration_of_phytoplankton_expressed_as_carbon_in_sea_water (PHYC)
- mole_concentration_of_dissolved_molecular_oxygen_in_sea_water (O2)
- mole_concentration_of_nitrate_in_sea_water (NO3)
- mole_concentration_of_phosphate_in_sea_water (PO4)
- surface_partial_pressure_of_carbon_dioxide_in_sea_water (SPCO2)
- sea_water_ph_reported_on_total_scale (PH)
- surface_downward_mass_flux_of_carbon_dioxide_expressed_as_carbon (FGCO2)
- net_primary_production_of_biomass_expressed_as_carbon_per_unit_volume_in_sea_water (PP)
- sea_water_alkalinity_expressed_as_mole_equivalent ()
- mole_concentration_of_dissolved_inorganic_carbon_in_sea_water (DIC)

In situ BGC observations are used in BS MFC for validation. These are oxygen (profiles and near real time surface data), chlorophyll (also in near real time), nitrate (including profiles), phosphate (including profiles), ammonium, silicate, Dissolved Inorganic Carbon, pH and Alkalinity, pC02, air-sea CO2 Flux. *Figure 7* shows that after 1995 the amount of data available in international databases have sharply decreased. A substantial part of the validation efforts has been targeted towards the assessment of the ability of the model to simulate oxygen. Significant BGC data gaps are found after 2010.

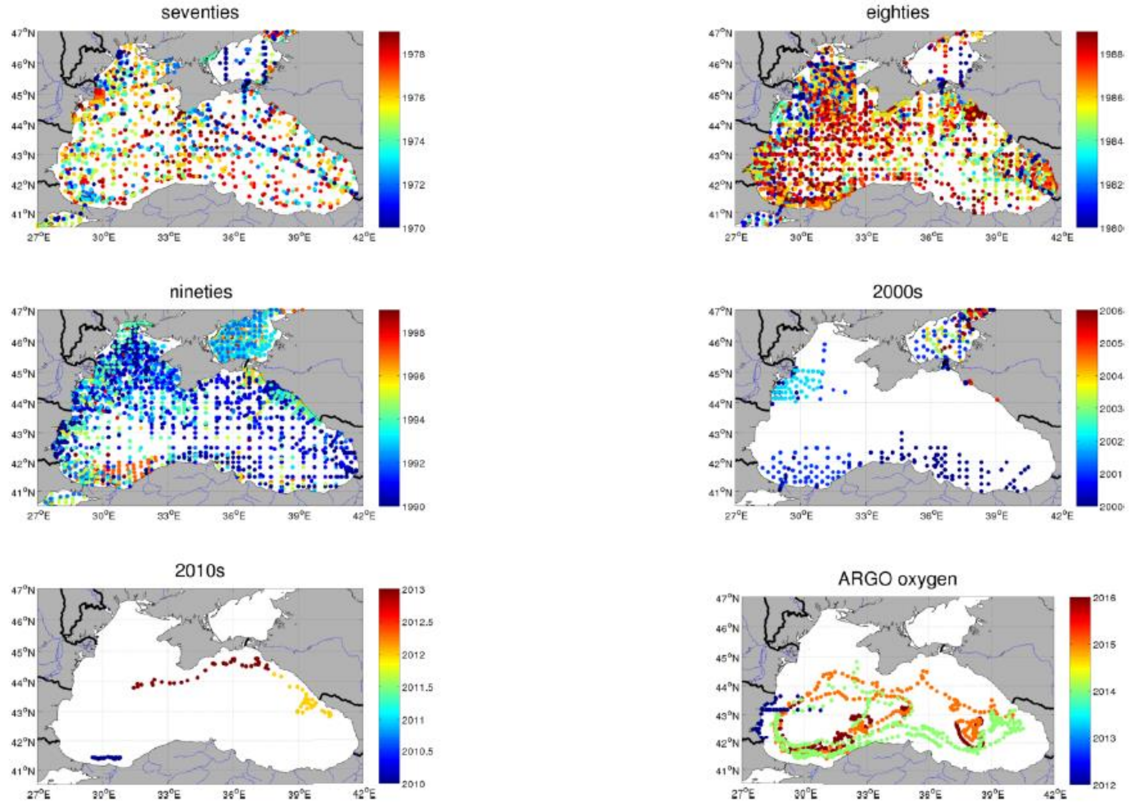

Figure 7. Distribution of data points per decade available in the NATO and WOD data base (the largest data bases in the Black Sea). Only sites where one of the modelled variables (oxygen, NH4, NO3/NO2, POC, PON, Si, TOC, and chlorophyll) has been collected are represented. The last plot shows the position of the Argo floats where oxygen has been collected. The different colors represent the year of sampling. ARGO floats are represented up to 2015 included. (Source : https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-BS-QUID-007-005.pdf)

Future requirements

According to the existing validation practice, there exist significant gaps on BGC observations in most of the Black Sea, especially after the 1990s (see *Figure 7*). Regarding the future requirements for the coastal ocean, requested variables are the ones associated with the determination of 1) optical properties: high quality chlorophyll (HPLC), radiometric data (IOPs and AOPs), turbidity (remote sensing validation data is currently scarce), SPM (with distinction between organic and mineral); 2) carbonate content: pH, alkalinity; 3) oxygen, inorganic nutrients; 4) fluxes at the interfaces: benthic-pelagic fluxes of oxygen, inorganic nutrients, DIC, and river nutrients inputs. These are variables that are labeled critical for coastal regions, more specifically for shallow and river-influenced regions.

Synthesis of results

Current status of the use of BGC variables

The current status of the use of BGC variables in the different regional CMEMS MFCs were assessed based on the publicly available Quality Information Documents (QUIDs). The current status is divided in two parts based on the purpose of the use in situ data: for data assimilation and for validation.

Data Assimilation

Table 1 and *Figure 8* provide an overview of BGC in-situ data being currently used for data assimilation in the different regional CMEMS MFCs. We can observe that currently only three MFCs use BGC in-situ data for data assimilation. BGC in-situ data being used for data assimilation includes chlorophyll-a concentration, as well as dissolved oxygen and nutrient (phosphate, nitrate, ammonium) profiles.

Table 1. BGC in-situ data currently used for data assimilation in the CMEMS MFCs.

	REGIONAL CMEMS MFCs					
BGC variables	**ARC MFC**	**BAL MFC**	**NWS MFC**	**IBI MFC**	**MED MFC**	**BS MFC**
Chlorophyll					X	
Phosphate	X	X			X	
Nitrate	X	X			X	
Ammonium		X				
Dissolved oxygen		X			X	

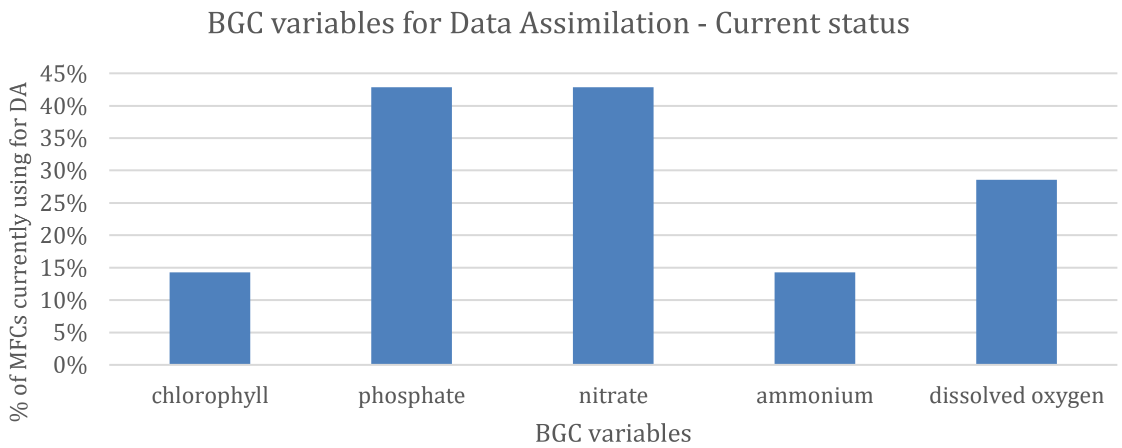

Figure 8. Overview of the BGC in-situ data currently used for data assimilation

Validation

BGC in-situ data currently used for validation in the regional CMEMS MFCs are shown in *Table 2* and *Figure 9*. All MFCs use chlorophyll-a, phosphate and nitrate observations, and almost all of them use dissolved oxygen and pH measurements, except the Arctic MFC.

Table 2. BGC variables currently validated with in-situ data in the CMEMS MFCs

	REGIONAL CMEMS MFCs					
BGC variables	**ARC MFC**	**BAL MFC**	**NWS MFC**	**IBI MFC**	**MED MFC**	**BS MFC**
chlorophyll	X	X	X	X	X	X
net primary production				X	X	
phytoplankton carbon concentration						
phosphate	X	X	X	X	X	X
nitrate	X	X	X	X	X	X
ammonium		X				X
silicate	X			X		X
dissolved oxygen		X	X	X	X	X
pH		X	X	X	X	X
pCO2		X	X	X	X	X
DIC				X	X	X
alkalinity					X	X
Secchi depth		X				
Light attenuation			X			
Euphotic layer depth				X		
CO2 air-sea flux					X	X

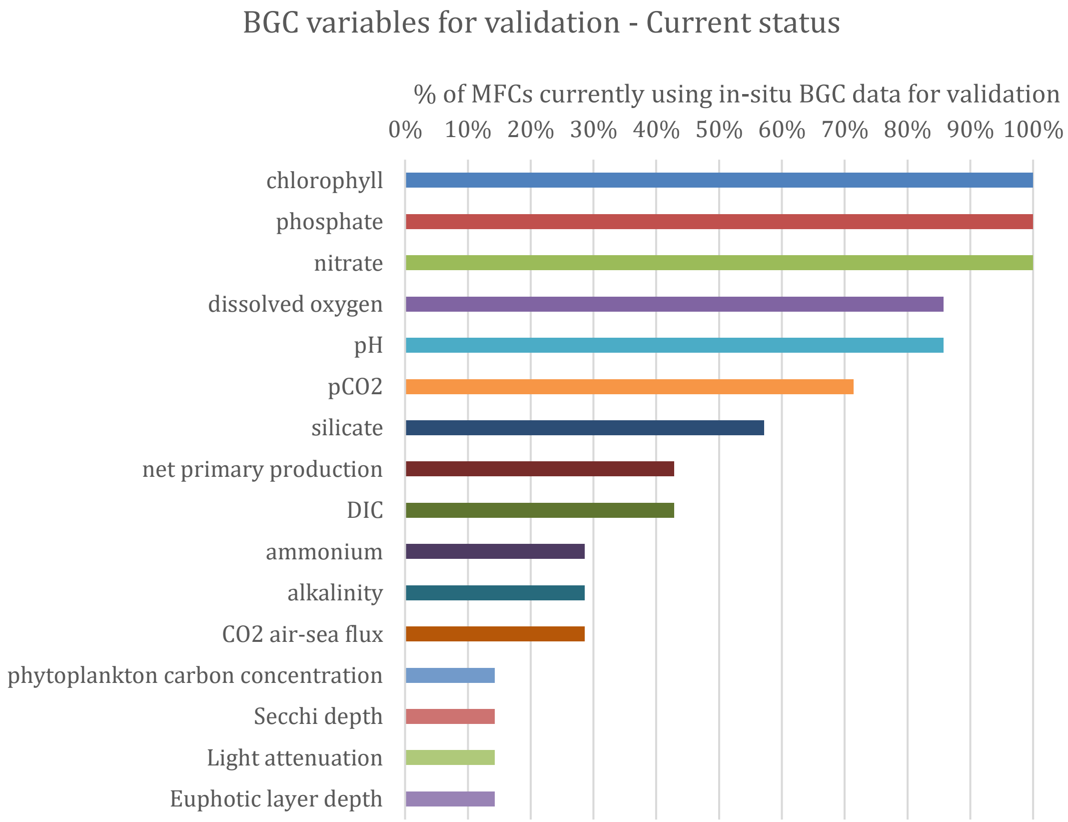

Figure 9. Overview of the BGC variables currently validated with in-situ data by the MFCs

Future requirements from the MFCs

After establishing the current status of BGC in-situ data use, the future requirements of CMEMS MFCs are summarized here. The future requirements were obtained from a survey launched to the MFCs (see Appendix). All MFCs filled in the survey, except the BS MFC, which communicated future requirements textually, not in the predefined tabular format. Since their future requirements were expressed in terms of variables, it was assumed that the purpose of use is calibration/validation as no indication of data assimilation or algorithm development was given. The future requirements were grouped based on the purpose of use (e.g. validation or data assimilation), and the data characteristics (spatio-temporal coverage and accuracy).

Purpose of use of the BGC in-situ data by the MFCs

In line with the current BGC in-situ data use, CMEMS MFCs indicate that in the future they will require chlorophyll-a, dissolved oxygen, and dissolved nutrient in-situ measurements the most, considering all purposes of use: algorithm development, validation and data assimilation (see Figure 10). It is important to note that **turbidity and phytoplankton biomass are among the BGC in-situ variables requested by MFCs, which are currently not being used, or being used only to a limited extent. This may indicate that MFCs aim to further develop modelling systems and improve prediction skills for turbidity and phytoplankton biomass simulations. Thus, providing in-situ observations of these variables may be an enabling factor towards improved BGC modelling capacity.**

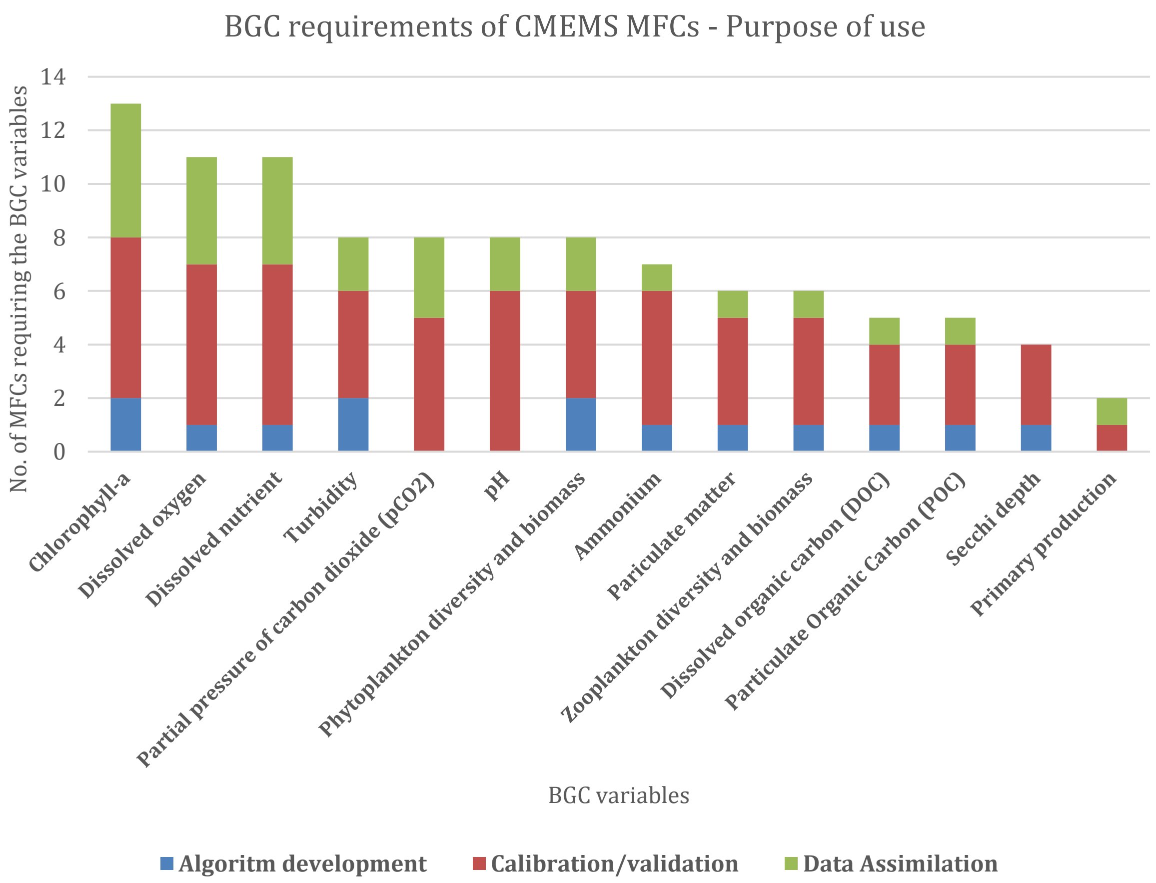

Figure 10. In-situ BGC requirements of CMEMS MFCs for all purposes

Looking at the different purposes of use separately (*Table 3*, *Table 4*, *Table 5*, *Figure 11*, *Figure 12*, *Figure 13*), it can be observed that mostly in-situ BGC data is needed for model calibration and validation, followed by data assimilation and algorithm development. For data assimilation currently only the Baltic, Mediterranean and Arctic MFCs are using BGC data other than chlorophyll-a. The future requirements, however, indicate that the other MFCs also wish to improve the data assimilation systems. All MFCs in the survey requested chlorophyll-a, and almost all of them require dissolved oxygen and dissolved nutrients.

Table 3. Future BGC in-situ data requirements in the CMEMS MFCs for algorithm development

	REGIONAL CMEMS MFC					
BGC variables	**ARC MFC**	**BAL MFC**	**NWS MFC**	**IBI MFC**	**MED MFC**	**BS MFC***
Turbidity			x	x		
Secchi depth				x		
Chlorophyll-a			x	x		
Dissolved oxygen			x			
Dissolved nutrient			x			
Ammonium			x			
Particulate matter			x			
Dissolved organic carbon (DOC)			x			
Particulate Organic Carbon (POC)			x			
Phytoplankton diversity and biomass		x	x			
Zooplankton diversity and biomass			x			

** The BS MFC did not specify the purpose of use for the future variable requirements, therefore calibration/validation was assumed*

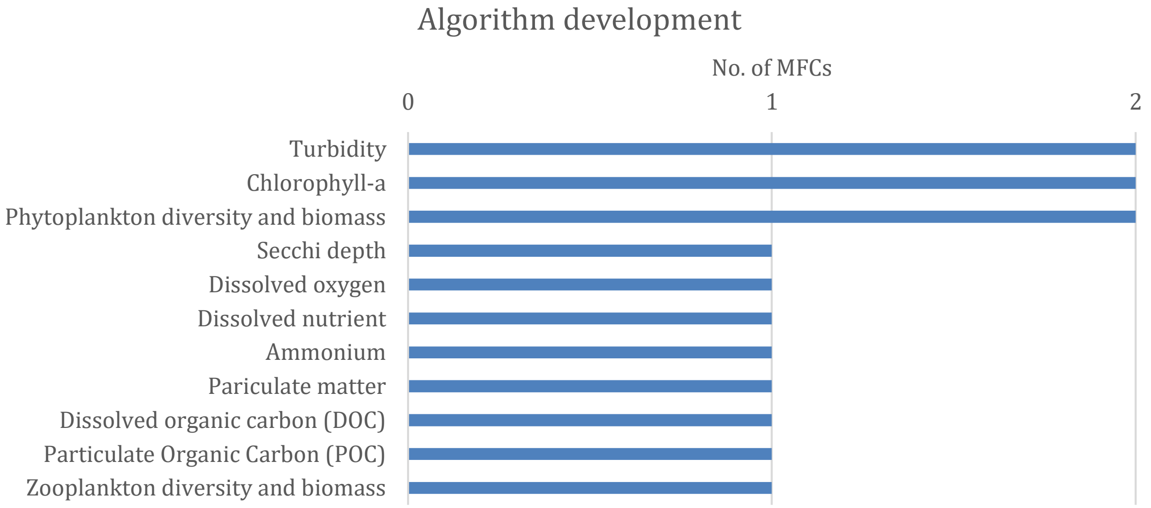

Figure 11. Number of MFCs requesting the BGC variable for algorithm development purposes.

Table 4. Future BGC in-situ data requirements in the CMEMS MFCs for model calibration and validation

	REGIONAL CMEMS MFC					
BGC variables	**ARC MFC**	**BAL MFC**	**NWS MFC**	**IBI MFC**	**MED MFC**	**BS MFC***
Turbidity	x	x	x			x
Secchi depth		x	x			x
Chlorophyll-a	x	x	x	x	x	x
Dissolved oxygen	x	x	x	x	x	x
Dissolved nutrient	x	x	x	x	x	x
Ammonium		x	x	x	x	x
Partial pressure of carbon dioxide (pCO2)	x	x	x	x	x	
Particulate matter			x	x	x	x
Dissolved organic carbon (DOC)			x	x	x	
Particulate Organic Carbon (POC)			x	x	x	
pH	x	x	x	x	x	x
Phytoplankton diversity and biomass	x		x	x	x	
Zooplankton diversity and biomass	x		x	x	x	
Primary production	x					

** The BS MFC did not specify the purpose of use for the future variable requirements, therefore calibration/validation was assumed*

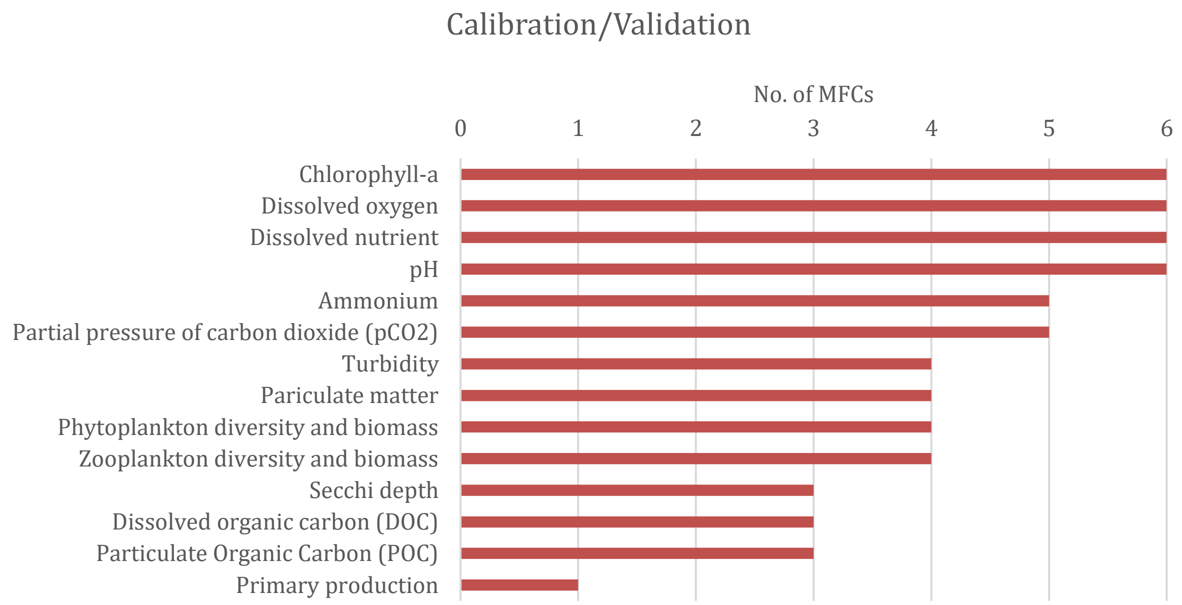

Figure 12. Number of MFCs requesting the BGC variable for calibration/validation purposes.

Table 5. Future BGC in-situ data requirements in the CMEMS MFCs for data assimilation

	REGIONAL CMEMS MFC					
BGC variables	**ARC MFC**	**BAL MFC**	**NWS MFC**	**IBI MFC**	**MED MFC**	**BS MFC***
Turbidity		x	x			
Secchi depth						
Chlorophyll -a	x	x	x	x	x	
Dissolved oxygen	x	x	x		x	
Dissolved nutrient	x		x	x	x	
Ammonium			x			
Partial pressure of carbon dioxide (pCO2)	x		x	x		
Particulate matter			x			
Dissolved organic			x			

carbon (DOC)			
Particulate Organic Carbon (POC)		x	
pH	x	x	
Phytoplankton diversity and biomass		x	x
Zooplankton diversity and biomass		x	
Primary production	x		

** The BS MFC did not specify the purpose of use for the future variable requirements, therefore calibration/validation was assumed*

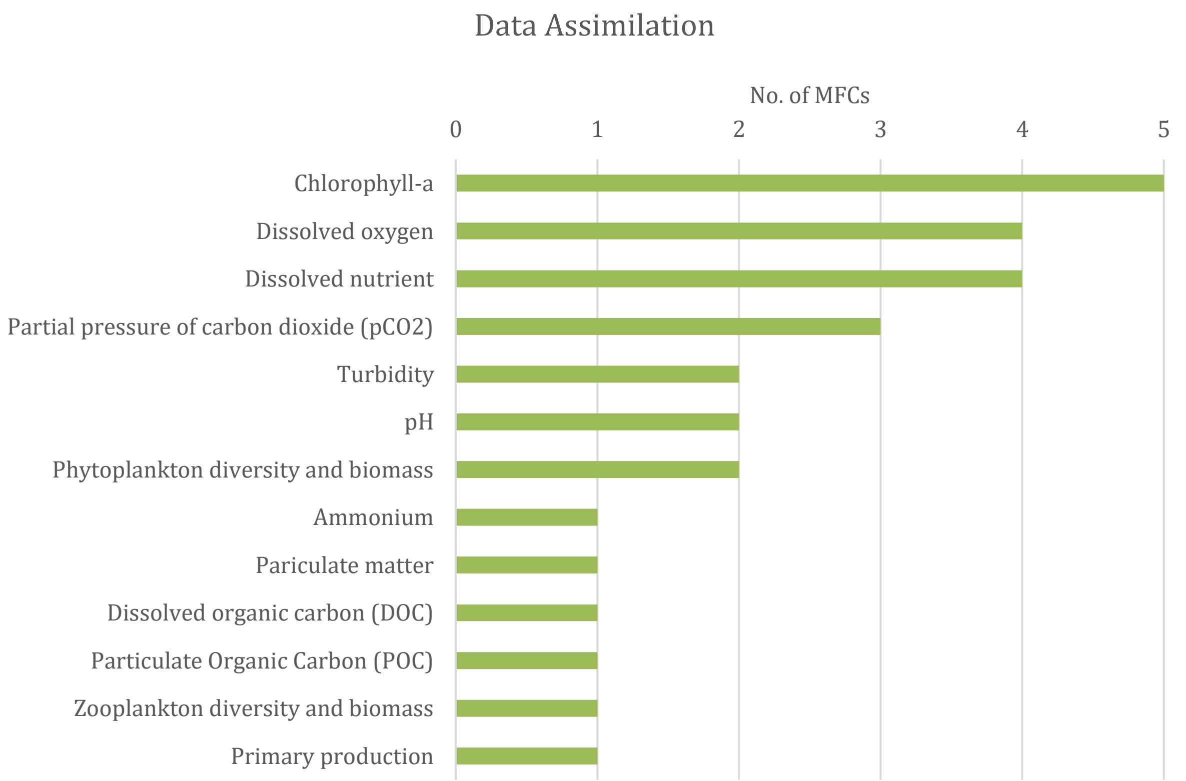

Figure 13. Number of MFCs requesting the BGC variable for data assimilation purposes.

Data Characteristics requirements

The requirements of BGC in-situ data expressed by the MFCs are further analysed in terms of spatial and temporal characteristics. **Regarding the spatial resolution and coverage, the requirements are largely heterogenous.** In the MED MFC, one station is required every 10 km of the coast for the entire Mediterranean coastline or near river outlets (zone of influence) e.g., Adriatic Sea, Aegean, France and Spain coastlines. Such spatial frequency seems rather hard to achieve given the current Mediterranean in-situ data coverage. Nevertheless, some Mediterranean coastlines already have relatively dense coverage such as the coastal observation network of ISPRA (>300 stations) (see *Figure 6*). Future effort will also need to focus on less monitored areas such as the Arctic. In the Arctic ocean, profiles are required every 100 km with at least 50 locations for the variables that are used for model validation and data assimilation. In other regional seas where more BGC data is currently available, there were no additional requirements stated apart from the fact that BGC in-situ data should cover parts of the Baltic and the North West Shelf region. In the Iberian area the spatial coverage requirements are the continental shelf and the mouths of major rivers for all BGC variables. **In general, existing BGC observations are not focused on the coastal areas but target the whole basin.** This is partially due to existing BGC data sources, such as BGC Argo floats or Ferry boxes, that cover more open ocean rather than coastal stations. **Improving and validating the forecasting skills of CMEMS MFCs in coastal areas will therefore require concentrated effort on the coast. Consequently, more BGC in-situ data sources are needed in the coastal area.**

Although data accuracy requirements were asked in the survey, MFC operators found it hard to express and quantify their requirements. **It was expressed that in general bias and uncertainty estimates are very helpful, but an absolute accuracy metric is less relevant.** Only the ARC MFC expressed a quantitative accuracy requirement of 10% on the chlorophyll-a data. In addition, the IBI MFC requires the best possible accuracy, indicating that data accuracy is indeed important.

Considering the **temporal aspects**, two type of requirements where assessed: update frequency and timeliness. The required update frequency is again highly variable depending on the purpose of the data use and the MFC. **Figure 14 shows that all BGC variables are required on a weekly basis (at least by one MFC) to allow product monitoring**. **The majority of BGC in-situ observations are also required with a rather sparse update frequency (monthly and six-monthly) to serve interim production and reanalysis purposes.** Furthermore, several of the considered BGC variables (chlorophyll-a, dissolved nutrients, ammonium, pCO2, pH, phytoplankton biomass) are required with a daily update frequency, and some are also required with a 6-hourly frequency (turbidity, chlorophyll-a, dissolved oxygen). The daily and 6 hourly update frequencies are most probably requested to meet operational data assimilation purposes. From *Figure 15* we can also observe that most variables are required in different update frequencies, except primary production which is only required weekly. The commonly used chlorophyll-a, dissolved oxygen, dissolved nutrient, ammonium, pCO2and pH variables are required at four update frequencies: 6 hourly, weekly, monthly, and 6 monthly.

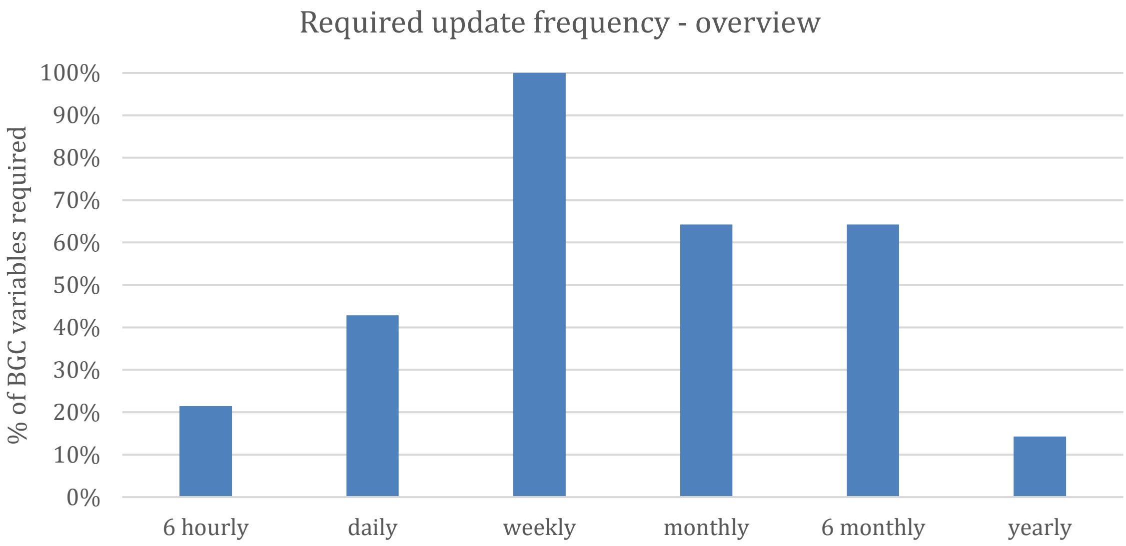

Figure 14. Required update frequency of BGC in-situ observations. The figure depicts the % of BGC variables required by at least one MFC.

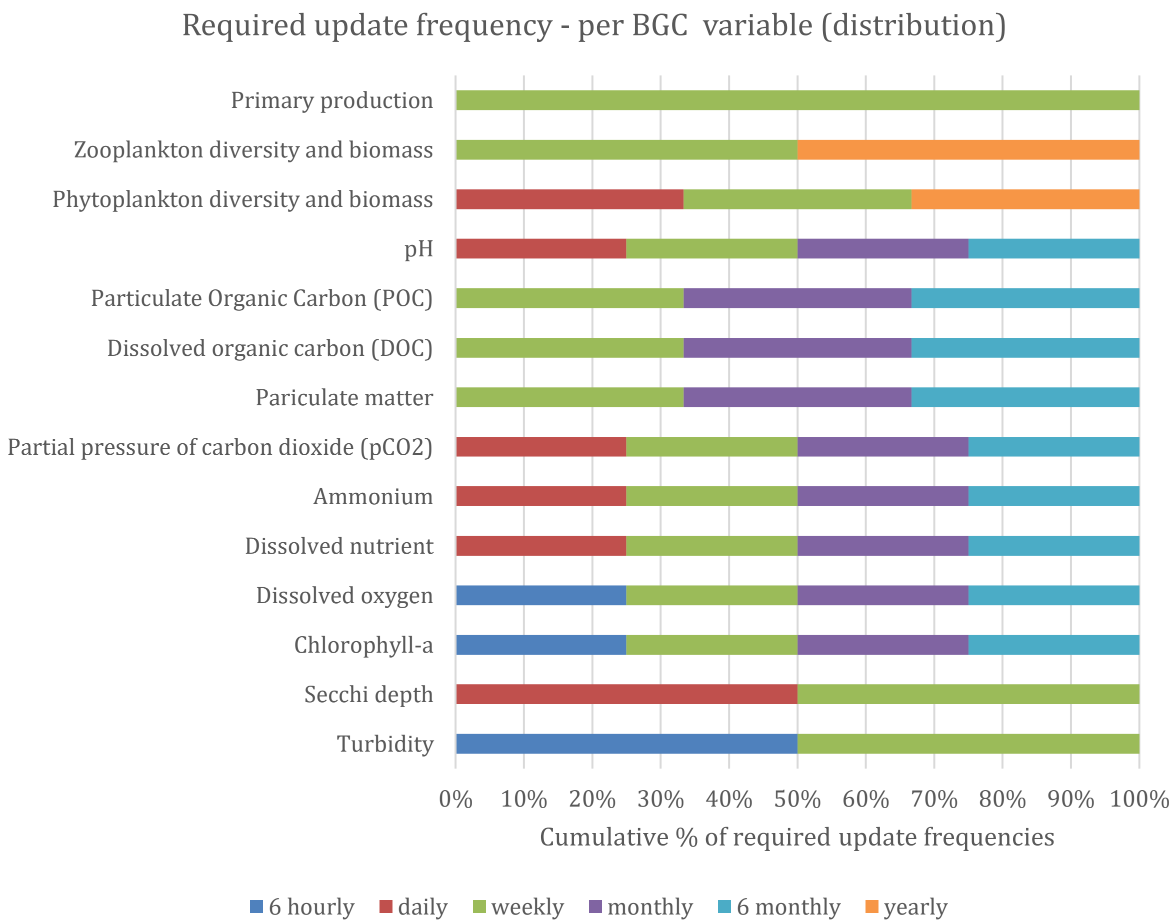

Figure 15. Required update frequency per BGC variable (distribution) showing the percentage that each update frequency contributes to the total (100%).

Finally, the requirements on the timeliness of data delivery are explored. In this respect, MFCs expressed various requirements which were then classified into four main groups: near real time (NRT), delayed mode (e.g., weekly), 3-6 months, or no requirement. *Figure 16* shows the requirements of CMEMS MFCs for each BGC variable. It is apparent that BGC variables are often required with different timeliness depending on the different purposes of use. For instance, chlorophyll-a is required with NRT, weekly, and 3-6 monthly timeliness as it serves operational data assimilation purposes, product monitoring, but also used for interim and reanalysis production. It should be noted that most (all) BGC in-situ data is required with 3-6 monthly timeliness but many are also required near-real time (by at least one MFC). The IBI MFC also expressed time coverage requirement per BGC variable. **The IBI MFC requires data throughout the year for dissolved oxygen, dissolved nutrients, ammonium, pCO2, Pariculate matter, DOC, POC and pH. On the other hand, spring and autumn data coverage is the most important for optical and biological variables: turbidity, secchi depth, chlorophyll-a, phytoplankton diversity and biomass, zooplankton diversity and biomass.**

Since the near real time (NRT) delivery requirement is the hardest to meet, we further investigated which BGC variables are required in NRT mode. It can be observed in *Figure 17* that chlorophyll-a is required the most in near real time mode. Moreover, half of the MFCs require dissolved oxygen, pCO2, particulate matter, and pH in NRT delivery mode. **Consequently, facilitating the requested near real time delivery for chlorophyll-a, dissolved oxygen, pCO2, particulate matter, and pH should be a priority.**

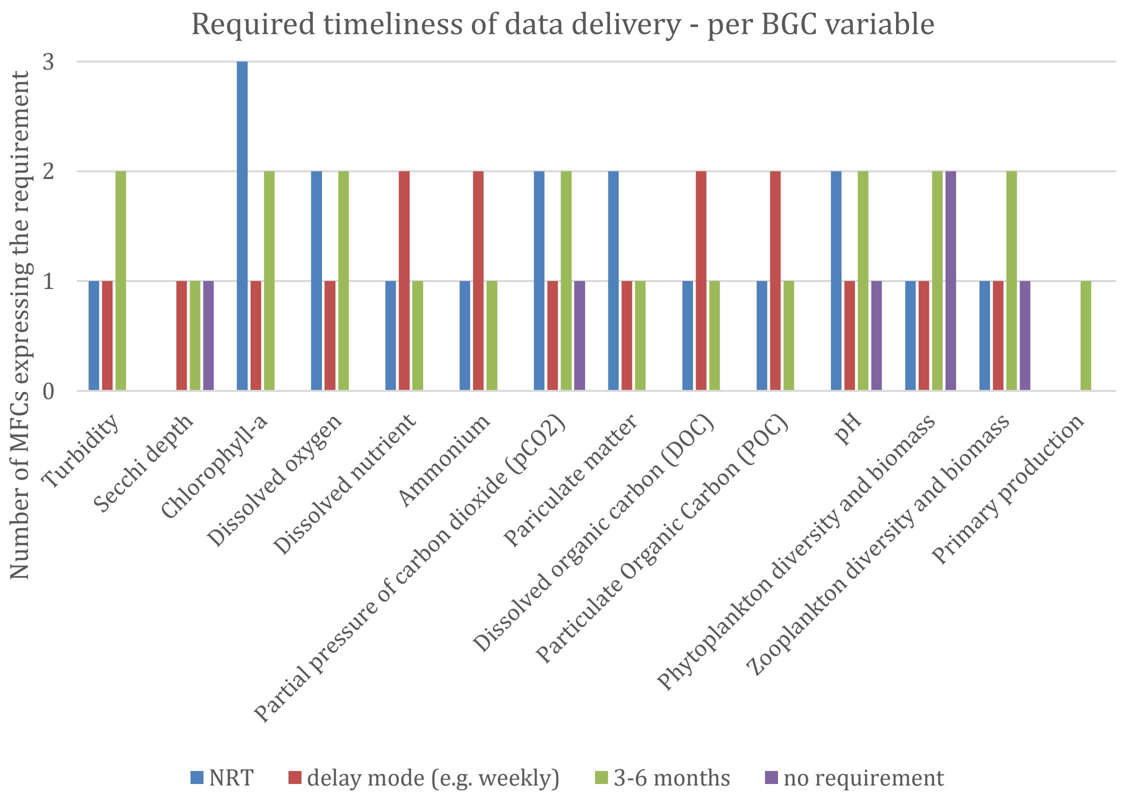

Figure 16. Required timeliness of data delivery - per BGC variable.

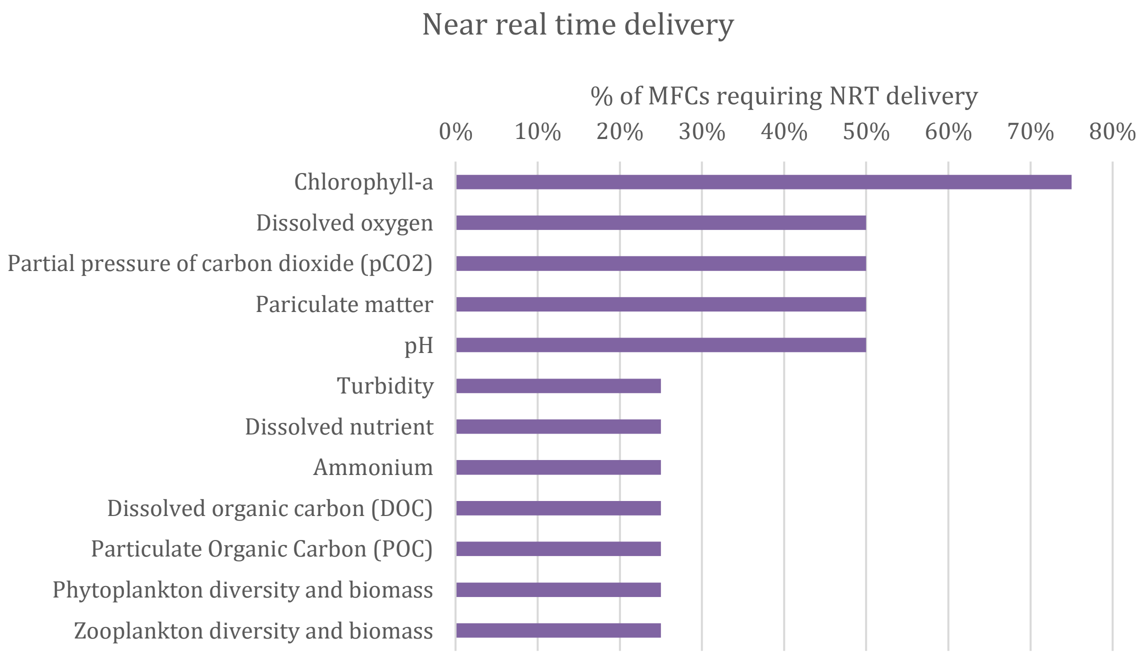

Figure 17. Requirement on near real time data delivery

Recommendations

Based on the requirements reported by the different MFCs, the following recommendations are formulated for the Copernicus in situ coordination team on Biogeochemistry Essential Ocean Variables in the coastal ocean:

- improving and validating the forecasting skills of CMEMS MFCs in coastal areas (fine scale processes and strong gradients) require concentrated effort on the coast, especially near mouths of major rivers. Consequently, more BGC in-situ data sources are needed in the coastal area;
- providing in-situ observations of turbidity and phytoplankton biomass, which are currently not used or only used to a limited extent, may be an enabling factor towards improved BGC modelling capacity in the coastal regions;
- future requirements indicate that the MFCs wish to improve their data assimilation systems including chlorophyll-a, dissolved oxygen, pCO2, particulate matter, and pH. Measuring these BGC variables should be prioritized;
- BGC in-situ observations are mostly required with weekly update frequency for product monitoring or even less frequently (monthly and six-monthly) to serve interim production and reanalysis purposes. Nevertheless, improved timeliness in data delivery is required to ensure data assimilation needs for coastal models, prioritizing near real time delivery (less than one day) for chlorophyll-a, dissolved oxygen, pCO2, particulate matter, and pH (see recommendation above);
- in the Iberian-Biscay-Irish MFCs optical and biological in-situ observations, such as turbidity, secchi depth, chlorophyll-a, phytoplankton and zooplankton, are most important during spring and autumn (the biologically most active periods). An increased number of such observations in these periods is required;
- the heterogenous spatial resolution and coverage requirements of MFCs call for regional (covering entire basins) and local (e.g., river outlets) monitoring programmes to fill the existing data gaps. New mechanisms need to be set up between the EU and member states to address the major gaps in coastal biogeochemical observations in areas of interest (e.g., Artic region or Black Sea);
- bias and uncertainty estimates of BGC in-situ data should be provided to CMEMS MFCs.

Appendix: Future BGC Requirements Survey template

The following table was circulated among CMEMS MFCs to Identifying and creating a list of user requirements (resolution in time and space, quality, timeliness) for Biogeochemistry (BGC) Essential Ocean Variables (EOVs) data in the coastal zone for Copernicus.

Variables*	Future requirements on BioGeoChemical data								
	Purpose of use (*Please tick where appropriate*)				Data characteristics (*Please provide numeric or textual input*)				
	Algorithm development	Calibration/validation	Data assimilation	Other (pls. specify)	No. of stations	Area covered	Update frequency	Timeliness	Accuracy
Turbidity									
Secchi depth									
Chlorophyll-a									
Dissolved oxygen									
Dissolved nutrient									
Ammonium									

Partial pressure of carbon dioxide (pCO2)									
Pariculate matter									
Dissolved organic carbon (DOC)									
Particulate Organic Carbon (POC)									
pH									
Phytoplankton diversity and biomass									
Zooplankton diversity and biomass									

***Note: Please amend the table if you feel that variables are missing.**