Using machine learning and BGC-Argo floats to assess biogeochemical models and optimize observing system design

Alexandre Mignot[1], Hervé Claustre[2,3], Gianpiero Cossarini[4], Fabrizio D'Ortenzio[2,3], Elodie Gutknecht[1], Julien Lamouroux[1], Paolo Lazzari[4], Coralie Perruche[1], Stefano Salon[4], Raphaëlle Sauzède[3], Vincent Taillandier[2,3], Anna Teruzzi[4]

[1]Mercator Océan International, Toulouse, France
[2]Laboratoire d'Océanographie de Villefranche-sur-Mer, Villefranche-sur-Mer, CNRS and Sorbonne Université, 06230 Villefranche-sur-Mer, France
[3]Institut de la Mer de Villefranche, CNRS and Sorbonne Université, 06230 Villefranche-sur-Mer, France
[4]National Institute of Oceanography and Applied Geophysics - OGS, Trieste, Italy

Corresponding author: Alexandre Mignot, amignot@mercator-ocean.fr

Numerical models of ocean biogeochemistry are becoming major tools to detect and predict the impact of climate change on marine resources and monitor ocean health. However, with the continuous improvement in model structure and spatial resolution, incorporation of these additional degrees of freedom into fidelity assessment has become increasingly challenging. Here, we propose a new method to inform about the model predictive skill in a concise way. The method is based on the conjoint use of a K-means clustering technique, assessment metrics and BGC-Argo observations. The K-means algorithm and the assessment metrics reduce the number of model data points to be evaluated. The metrics evaluate either the model state accuracy or the skill of the model in capturing emergent properties, such as the Deep Chlorophyll Maximums and Oxygen Minimum Zones. The use of BGC-Argo observations as the sole evaluation data set ensures the accuracy of the data as it is a homogenous data set with strict sampling methodologies and data quality control procedures. The method is applied to the global ocean biogeochemical analysis and forecasting system of the Copernicus Marine Service. The model performance is evaluated using the model efficiency statistical score that compares the model-observations misfit with the variability of the observations, and thus objectively quantifies whether the model outperforms the BGC-Argo climatology.

We show that, overall, the model surpasses the BGC-Argo climatology in predicting pH, dissolved inorganic carbon, alkalinity, oxygen, nitrate, and phosphate in the mesopelagic and the mixed layers, as well as silicate in the mesopelagic layer. However, there are still areas for improvement in reducing the model-data misfit for certain variables such as silicate, pH, and the partial pressure of CO_2 in the mixed layer, as well as chlorophyll-a related, Oxygen Minimum Zones-related and particulate organic carbon metrics. The method proposed here is also helpful to inform the design of the BGC-Argo network, in particular, the regions where BGC-Argo observations should be enhanced to improve the model accuracy through the assimilation of BGC-Argo data or process-oriented assessment studies. We strongly recommend increasing the number of observations in the Arctic region, while maintaining the already high-density of observations in the Southern Oceans. The model error in these regions is only slightly less than the variability observed in BGC-Argo measurements. Our study illustrates how the synergic use of modelling and BGC-Argo data can both inform about the performance of models and the design of observing systems.

1. Introduction

Since pre-industrial times, the ocean has taken ~26 % of the total anthropogenic CO_2 emission (Friedlingstein et al., 2022) leading to dramatic change in the ocean's biogeochemical (BGC) cycles, such as ocean acidification (Iida et al., 2020). Moreover, deoxygenation (Breitburg et al., 2018) and change in the biological carbon pump are now manifesting globally (Capuzzo et al., 2018; Osman et al., 2019; Roxy et al., 2016). Together with plastic pollution (Eriksen et al., 2014) and an increase in fisheries pressure (Crowder et al., 2008), major changes are therefore occurring in marine ecosystems at the global scale. In order to contextualize monitoring of ongoing changes, derive climate projections and develop better mitigation strategies, realistic numerical simulations of the oceans' BGC state are required.

Numerical models of ocean biogeochemistry represent a prime tool to address these issues because they produce three dimensional estimates of a large number of chemical and biological variables that are dynamically consistent with the ocean circulation (Fennel et al., 2019). They can assess past and current states of the BGC ocean, produce short-term to seasonal forecasts as well as climate projections. However, these models are far from being flawless, mostly because there are still huge knowledge gaps in the understanding of key BGC processes and, as a result, the mathematical functions that describe BGC fluxes, and ecosystems dynamics are too simplistic (Schartau et al., 2017). For instance, most models do not include a radiative component for the penetration of solar radiation in the ocean. It has been nevertheless shown that coupling such a component with a BGC model improves the representation of the dynamics of phytoplankton in the lower euphotic zone (Dutkiewicz et al., 2015; Álvarez et al., 2022). Additionally, the parameterization of the mathematical functions generally results from laboratory experiments on a few representative species and may not be suitable for extrapolation to ocean simulations that need to represent the large range of organisms present in oceanic ecosystems (Schartau et al., 2017; Ward et al., 2010). Furthermore, the assimilation of physical data in coupled physical-BGC models that improves the physical ocean state can paradoxically degrade the simulation of the BGC state of the ocean (Fennel et al., 2019; Park et al., 2018; Gasparin et al., 2021). A rigorous assessment of BGC models is thus essential to test their predictive skills, their ability to reproduce BGC processes and estimate confidence intervals on model predictions (Doney et al., 2009; Stow et al., 2009).

However, the evaluation of BGC models is limited by the availability of data. It relies principally on a combination of different data sets from satellite (such as chlorophyll-*a* concentration), cruises observations, permanent oceanic stations from large databases such as the World Ocean Database (e.g., Doney et al., 2009; Dutkiewicz et al., 2015; Lazzari et al., 2012, 2016; Lynch et al., 2009; Séférian et al., 2013; Stow et al., 2009). All these datasets have neither a sufficient vertical or temporal resolution, nor a synoptic view, nor provide all variables necessary to evaluate how models represent climate-relevant processes such as the air-sea CO_2 fluxes, the biological carbon pump, ocean acidification or deoxygenation.

In 2016, the Biogeochemical-Argo (BGC-Argo) program was launched with the goal to operate a global array of 1000 BGC-Argo floats equipped with oxygen (O_2), chlorophyll *a* (Chl*a*) and nitrate (NO_3) concentrations, particulate backscattering (b_{bp}), pH and downwelling irradiance sensors (Biogeochemical-Argo Planning Group, 2016; Claustre et al., 2020). Although the planned number of 1000 floats has not been reached yet, the BGC-Argo program has already provided a large number of quality-controlled vertical profiles of O_2, Chl*a*, NO_3, b_{bp}, and pH (Fig. 1). With respect to O_2, Chl*a*, NO_3, and b_{bp}, the North Atlantic and the Southern Ocean are reasonably well sampled whereas pH is well sampled only in the Southern Ocean. At the regional scale, the Mediterranean Sea is also fairly well sampled by BGC-Argo floats (Salon et al., 2019; Terzić et al., 2019; D'Ortenzio et al., 2020). However, there are still large under-sampled areas like the Arctic Ocean, subtropical gyres and the sub-polar North Pacific. Thanks to machine learning based methods (Bittig et al., 2018; Sauzède et al., 2017), floats equipped with O_2 sensors can be additionally used to derive vertical profiles of NO_3, phosphate (PO_4), silicate (Si), alkalinity (Alk), dissolved inorganic carbon (DIC), pH and pCO_2.

The BGC-Argo data set represents a significant improvement for the assessment of models compared to large databases such as the World Ocean Database (Boyer et al., 2013) or the Copernicus Marine Service in situ dataset (European Union-Copernicus Marine Service, 2015). Large databases are composed of data collected from various instrument types with heterogenous data sampling methodologies. Therefore, for a given variable, the accuracy numbers are not the same and change depending on the instrument type (European Union-Copernicus Marine Service, 2019). Consequently, this affects the overall accuracy over time due to the changing proportion of instrument types over the years. On the other hand, the BGC-Argo data set is a homogenous data set with strict and uniform sampling methodologies and data Quality-Control (QC) procedures. As a result, the BGC-Argo data set has a satisfactory level of accuracy, which remains stable over time (Johnson et al., 2017; Mignot et al., 2019). Moreover, the number of quality-controlled observations collected every year by the BGC-Argo fleet is now greater than any other data set (Claustre et al., 2020). Using the BGC-Argo data set as the single evaluation data set is therefore a way to ensure consistent accuracy.

The BGC-Argo floats provide multivariate observations at high vertical and temporal resolutions and for long periods of time providing nearly continuous time series of the vertical distribution of several biogeochemical variables. This is not possible with discrete, univariate vertical samplings provided by cruise cast *in situ* measurements or from climatological values derived from the World Ocean Atlas. All these specificities overcome the limitations of previous datasets, especially with respect to their univariate nature, as well as their limited vertical and temporal resolutions. This opens new perspectives for the evaluation of BGC models (Gutknecht et al., 2019; Salon et al., 2019; Terzić et al., 2019).

The development of BGC models, coupled with the ongoing increase in spatial and vertical resolutions, has resulted in a significant rise in the volume of model outputs. Simplification techniques are therefore required to provide decipherable information on model predictive skill. Allen et al. (2007) proposed a methodology for reducing the spatial dimensions in model assessment exercises, thereby providing concise information about the model performance. They use an unsupervised learning algorithm to classify the southern North Sea into 5 coherent BGC regions based on modelled time series of temperature, NO_3, PO4, and Si concentrations. Then, they evaluated the predictive capabilities of the model in each BGC region (instead of each grid point), thus greatly reducing the number of points to be validated. An additional method for reducing the dimensions of model-data comparison is the use of assessment metrics (Hipsey et al., 2020; Russell et al., 2018). In particular, metrics such as depth-averaged state variables (e.g., mixed layer averaged Chl*a*, NO_3, O_2, etc…), mass fluxes and process rates (e.g., primary production or division rates), or emergent properties (e.g., Deep Chlorophyll Maximum (DCM), or Oxygen Minimum Zone (OMZ)) are particularly useful to reduce the number of model's vertical layers to be compared with the observations.

The objectives of the present study are twofold. Our first aim is to propose a methodology that uses the BGC-Argo data set, an unsupervised learning algorithm and assessment metrics to simplify marine BGC model-data comparisons, and thus inform, in a concise way, about model performances. The second objective is to use this methodology to also identify ocean regions where the model-observations misfit is larger than the variability of the BGC-Argo data and thus inform the BGC-Argo observing system of regions that should be better sampled. The first step of the method consists in defining 23 assessment metrics that are used both to construct the BGC regions and then to compare the model outputs with the BGC-Argo data. Second, following the approach of Allen et al. (2007), we use an unsupervised learning algorithm, specifically a K-means clustering technique, to classify the global ocean into 8 coherent BGC regions based on the climatological modelled time series of the 23 assessments metrics. In the last step, the skill of the model in predicting the assessment metrics is evaluated in each BGC-region, using the model efficiency statistical score. Unlike other statistical metrics such the correlation coefficient, the bias or the root mean square difference, that does not quantify objectively whether the model performance is acceptable or not; the model efficiency calculates whether the model outperforms an observational climatology (Fennel et al., 2022). Finally, the method is implemented using the global ocean BGC analysis and forecasting system of the Copernicus Marine Service (European Union-Copernicus Marine Service, 2019).

The paper is organised as follows: section 2 presents the data sets used in the study. In section 3, we define the assessment metrics, and we detail the K-means algorithm as well as the model efficiency statistical score. In section 4, we present and discuss the results. Finally, section 5 concludes the study.

2. Data

a. BGC-Argo floats observations

The float data were downloaded from the Argo Coriolis Global Data Assembly Centre in France (ftp://ftp.ifremer.fr/argo, last accessed in January 2023). The CTD and trajectory data were quality controlled using the standard Argo protocol (Wong et al., 2015). The raw BGC signals were transformed to biogeochemical variables (i.e., O_2, Chl*a*, NO_3, b_{bp}, and pH) and quality-controlled according to international BGC-Argo protocols (Johnson et al., 2018b, a; Schmechtig et al., 2015, 2018; Thierry et al., 2018; Thierry and Bittig, 2018).

In the Argo data-system, the data are available in three data modes: "Real-Time", "Adjusted" and "Delayed" (Bittig et al., 2019). In the "Real-time" mode, the raw data are converted into state variables and an automatic quality-control is applied to "flag" gross outliers. In the "Adjusted" mode, the "Real-time" data receive a calibration adjustment in an automated manner. In the "Delayed" mode, the "Adjusted" data are adjusted and validated by a scientific expert. While the "Real-Time" and "Adjusted" data are considered acceptable for operational application (data assimilation), the "Delayed" mode is designed for scientific exploitation and represent the highest quality of data with the ultimate goal, when time-series with sufficient duration will have been acquired, to possibly extract climate-related trends (Bojinski et al., 2014). However, for some variables, only a limited fraction of data is accessible in "Delayed" mode. Consequently, for each variable, we selected the highest level of data modes, where at least 80 % of the data are available (see Table 1). Note that this criterion is not applied to O_2, where only "Delayed" mode data were selected in order to generate the pseudo-observations from CANYON-B neural network (see after). We removed data with missing location or time information and flagged as "Bad data" (flag =4). Depending on the parameter and the associated data mode, we also excluded data flagged as "potentially bad data" (flag=3) (see Table 1). Finally, it should be noted that the status of the different modes of adjustment for b_{bp} is still very inhomogeneous in the global BGC-Argo database. A quality control procedure in "Real-Time" has just been proposed to the Argo Data Management Team but is not yet operationally implemented in the database (Dall'Olmo et al. 2022). Since there is no current official consensus for the qualification of b_{bp} data we decided to use for this study all data modes but to remove the data that are flagged as "Bad data" (see details in Table 1).

Particulate Organic Carbon (POC) concentrations were derived from b_{bp} observations. First, three consecutive low-pass filters were applied on the vertical profiles of b_{bp} to remove spikes (Briggs et al., 2011): a 2-point running median followed by a 5-point running minimum and 5-point running maximum. Then, the filtered b_{bp} profiles were converted into POC (mgC m^{-3}) using a simplified version of the empirical POC/b_{bp} algorithm developed by Gali et al. (2022), i.e , for depths larger than the mixed layer depth (MLD):

$$\frac{POC}{b_{bp}} = c + a \cdot e^{-0.001 \cdot b \cdot (z - MLD)}, \qquad (1)$$

$$z > MLD,$$

where c is a constant deep value and, b, the slope of the exponential decrease, sets to 12010 mgC m^{-3} m and -6.57, respectively, as proposed by Gali et al. (2022). The global coefficient a, is set to 37990 mgC m^{-3} m to be consistent with a relationship, developed for global applications (i.e, POC= 38687.27* $b_{bp}{}^{0.95}$) (European Union-Copernicus Marine Service, 2020). In the Mixed Layer (ML), z is fixed at z = MLD, and the Eq. (1) simplifies to

$$\frac{POC}{b_{bp}} = c + a, \qquad (2)$$

$$z \le MLD.$$

Finally, we complemented the existing BGC-Argo dataset with pseudo-observations of NO_3, PO_4, Si, Alk, and DIC concentrations as well as pH and pCO_2 using the CANYON-B neural network (Bittig et al., 2018). CANYON-B estimates vertical profiles of nutrients as well as the carbonate system variables from concomitant measurements of float pressure, temperature, salinity, and O_2 qualified in "Delayed" mode together with the associated geolocalization and date of sampling. CANYON-B was trained and validated using the GLODAPv2 data set (Key et al., 2015). The CANYON-B estimates of NO_3 and pH were merged with measured values on the rationale that CANYON-B estimates have RMS errors (NO_3 = 0.7 µmol kg^{-1}, pH = 0.013) (Bittig et al., 2018) that are of the same order of magnitude as those of the BGC-Argo observations errors (NO_3 = 0.5 µmol kg^{-1}, pH = 0.07) (Mignot et al., 2019; Johnson et al., 2017) .

Finally, we verified that the RMS errors of BGC-Argo data (both measured and from CANYON-B estimates) are lower than the RMS difference between the model and BGC-Argo data, so that the comparison of simulated properties with the BGC-Argo data leads to a meaningful evaluation of the model performance. We believe it is reasonable to draw conclusions on the model uncertainty from BGC-Argo data as long as the BGC-Argo errors are much lower than the model-observations RMS difference.

b. Global Ocean BGC analysis and forecasting system of the Copernicus Marine Service

The global model simulation used in this study (see Appendix A.1) originates from the global ocean hydrodynamic-biogeochemical coupled system, based on NEMO-PISCES model, implemented and operated by Mercator Ocean for the Marine Service of the EU's earth observation programme Copernicus (CMEMS, 2020). The BGC component is constrained by the assimilation of satellite Chl*a* concentrations, and a climatological-damping is applied to nitrate, phosphate, oxygen, silicate - with World Ocean Atlas 2013 - to dissolved inorganic carbon and alkalinity – with GLODAPv2 climatology (Key et al., 2015) - and to dissolved organic carbon and iron - with a 4000-year PISCES climatological run. The BGC model is forced in offline mode by daily averages of ocean physics, sea ice and atmospheric conditions. The ocean physics and sea ice forcing come from the global ocean physics analysis and forecasting system at 1⁄12° (Lellouche et al., 2018) that assimilates along-track altimeter data, satellite Sea Surface Temperature and Sea-Ice Concentration, and *in situ* temperature and salinity vertical profiles. The BGC model has a 1/4° horizontal resolution, 50 vertical levels (with 22 levels in the upper 100 m, the vertical resolution is 1 m near the surface and decreases to 450 m resolution near the bottom).

We used daily outputs of Chl*a*, NO_3, PO_4, Si, O_2, pH, DIC and Alk, and weekly outputs of the wo size classes of phytoplankton, the small detrital particles and microzooplankton (resampled offline from weekly to daily frequency through constant interpolation) from 2009 to 2020. Note that the method of linear resampling, while artificially increasing the number of data, could potentially bias the statistical results, especially in regions with poor data coverage. As suggested by Gali et al. (2022), the POC concentration was computed offline by adding together the two size classes of phytoplankton, the small detrital particles and microzooplankton modelled by PISCES. This particular combination of phytoplanktonic and non-phytoplanktonic organisms has been shown to match the small POC observed by the floats. The partial pressures of CO_2 values were extrapolated in the mixed layer from the surface value estimated by the model. The Black Sea was not considered in the present analysis because the model solutions are of poor quality. Finally, the daily model outputs were collocated in time and space the closest to the BGC-Argo floats positions, and they were interpolated to the sampling depth of the float observations. The characteristics of the model are further detailed in the appendix.

3. Methods

a. Assessment metrics

In this section, we present 23 metrics used for the clustering of the ocean and for the assessment of the model simulation with BGC-Argo data. The metrics are associated with the carbonate chemistry, the biological carbon pump, and oxygen levels. Most of the metrics evaluate the model state accuracy through the comparison of simulated state variables with BGC-Argo observations depth-averaged in the mixed (hereinafter indicated with the subscript $_{mixed}$) and mesopelagic (hereinafter indicated with the subscript $_{meso}$) layers. This two-layer comparison between model and BGC-Argo data provides an indirect evaluation of the key processes and fluxes associated with the carbonate chemistry, biological carbon pump and oxygen levels in the mixed and mesopelagic layers. In addition, some of the metrics assess the skill of the model in capturing emergent properties, such as the nitracline, the DCMs and the OMZs. The metrics are described below and summarized in Table 2. The definition of the metrics is the same for the model and the BGC-Argo data. The MLD is computed, following De Boyer et al. (2004), as the depth at which the change in potential density from its value at 10 m exceeded 0.03 kg m^{-3}. Dall'Olmo and Mork (2014) define the mesopelagic layer as the region between the deeper of either the euphotic layer depth or the MLD, and a depth of 1000 meters. However, for ease of use, we adopt a simplified definition that considers the mesopelagic layer to be the region between the MLD and a depth of 1000 meters. To ensure the accuracy of the metrics calculation, we have checked the representation of the MLDs in the model. The model's MLDs closely match the observed data, as indicated by an overall mean square difference of approximately 30% of the total variance in the observations.

i. Carbonate chemistry

The uptake of ~26 % anthropogenic CO_2 by the global ocean (Friedlingstein et al., 2022) has altered the oceanic carbonate chemistry over the past few decades (Iida et al., 2020). Assessing how models correctly represent the oceanic carbonate chemistry is therefore critical if we aim to derive accurate climate projections on their future change. The classical variables for the study of carbonate chemistry are DIC, Alk, pH and pCO_2 (Williams and Follows, 2011). These variables are assessed in the mixed (DIC_{mixed}, Alk_{mixed}, pH_{mixed} and $pCO_{2\ mixed}$) and mesopelagic (DIC_{meso}, Alk_{meso}, pH_{meso}) layers. The partial pressure of CO_2 is only assessed in the mixed layer as the evaluation of $pCO_{2\ mixed}$ plays a critical role to assess the skill of BGC models to correctly represent the air-sea CO_2 flux.

ii. Biological carbon pump

The biological carbon pump is the transformation of nutrients and dissolved inorganic carbon into organic carbon in the upper part of the ocean through phytoplankton photosynthesis and the subsequent transfer of this organic material into the deep ocean. The functioning of this pump relies on key pools of nutrients and carbon as well as several processes that control mass fluxes between the pools. Changes in the biological carbon pump are now manifesting globally (Capuzzo et al., 2018; Osman et al., 2019; Roxy et al., 2016).

One way to indirectly evaluate the model's ability to accurately capture essential processes related to the biological carbon pump in the ocean's upper layer, such as primary production, respiration, and grazing, is to compare various ML pools [here the nutrients ($NO_{3\ mixed}$, $PO_{4\ mixed}$, Si_{mixed}), Chl_{mixed} and POC_{mixed}] with BGC-Argo observations. Similarly, the assessment of the mesopelagic nutrients, and POC concentration (hereinafter denoted $NO_{3\ meso}$, $PO_{4\ meso}$, Si_{meso}, and POC_{meso}) provides an indirect evaluation of the key mesopelagic layer processes, such as export production, respiration, etc.

In stratified systems, a DCM is formed at the base of the euphotic layer (Barbieux et al., 2019; Cullen, 2015; Letelier et al., 2004; Mignot et al., 2014, 2011). It has been suggested that the DCM plays a key role in the synthesis of organic carbon by phytoplankton (Macías et al., 2014). DCMs are therefore key features to be assessed in BGC models with respect to processes involved in the biological carbon pump such as the primary production. However, the DCM layer generally escapes detection by remote sensing. Furthermore, the DCM is also an emergent feature that develops in response to complex physical and biogeochemical interactions (Cullen, 2015). Thus, its evaluation provides critical information regarding the accuracy of the model in capturing complex patterns of key ecosystem processes. The depth and magnitude of DCM (H_{DCM} and Chl_{DCM}) are helpful metrics for the assessment of DCM dynamics. The depth of the DCM is calculated as the depth where the maximum of Chl*a* occurs in the profile with the criterion that H_{DCM} should be deeper than the MLD. The magnitude of the DCM corresponds to the Chl*a* value at H_{DCM}.

NO_3 is often depleted in the surface layers and is a limiting factor for phytoplankton growth in most oceanic regions. The vertical supply of NO_3 to the surface layers depends, among other factors, on the vertical gradient of NO_3 (the nitracline), and, in particular, on its depth (the nitracline depth) (Cermeno et al., 2008; Omand and Mahadevan, 2015). Therefore, the comparison of the simulated nitracline depth (H_{nit}) with BGC-Argo observations allows for an indirect assessment of the model performance in reproducing vertical fluxes of NO_3. Following previous studies (Cermeno et al., 2008; Lavigne et al., 2013; Richardson and Bendtsen, 2019), the depth of the nitracline is identified as the first depth where NO3 is detected. A detection threshold of 1 μmol kg^{-1} is used, which is an upper estimate of the accuracy of BGC-Argo NO3 data (Johnson et al., 2017; Mignot et al., 2019).

iii. Oxygen levels

Oxygen levels in the global and coastal waters have declined over the whole water column over the past decades (Schmidtko et al., 2017) and OMZs are expanding (Stramma et al., 2008). Assessing how models correctly represent ocean oxygen levels as well as the OMZs is therefore critical to monitor their change over time. Similar to the assessment of DCMs, evaluating Oxygen Minimum Zones (OMZs) provides insight into how the model represents emergent dynamics resulting from intricate physical and biogeochemical interactions (Paulmier and Ruiz-Pino, 2009). Oxygen levels are evaluated in the mixed (O_{2mixed}) and mesopelagic (O_{2meso}) layers. OMZs are defined as oceanic regions where O_2 levels are lower than 20 µmol kg^{-1} (Paulmier and Ruiz-Pino, 2009). OMZs are characterized by their depths (H_{O2min}) and their concentrations (O_{2min}).

b. Bioregionalization of the global ocean

In this study, we use the K-means clustering algorithm (Hartigan and Wong, 1979) to regionalize the ocean based on the modelled climatological monthly time series of the 23 metrics described previously. *T*he K-means clustering algorithm is an unsupervised machine learning technique that groups similar objects together in a way that maximizes similarity between objects within a group and minimizes similarity between objects in different groups. This clustering tool has been successfully used to classify marine BGC regions in different oceanic basins based on the seasonal cycle of satellite chlorophyll (Kheireddine et al., 2021; Mayot et al., 2016; Lacour et al., 2015; D'Ortenzio and d'Alcala, 2009) . The step-by-step methodology, used in this study, is described in the next section.

The first step in the analysis involved computing monthly climatological time series for the 23 metrics at each model grid cell. These time series were derived from the monthly climatological time series of state variables predicted by the model from 2009 to 2020. To account for the log-normal distribution and the wide range of values for metrics in units of Chla or POC, a log-10 transformation was applied to these metrics. Second, to take into consideration the 6-months shift in seasons between the northern and southern hemispheres, the dates for grid cells located in the Southern Hemisphere were shifted by 6 months (Bock et al., 2022). Third, to classify model grid cells based on the seasonality and amplitude of the 23 metrics, each metric was standardized by subtracting the global mean and dividing by the global standard deviation. This ensured that each metric had a mean of 0 and a standard deviation of 1, enabling comparison across metrics with different units. Fourth, to reduce the dimensionality of the data set, a principal component analysis was applied to the scaled data. Only the components that explain 99 % of the variance in the data set were kept, reducing thereby the dimensions of the data set by 85 %. A K-means clustering analysis was then performed on the resulting data set. Following Kheireddine et al. (2021), the number of clusters was determined based on a silhouette analysis (Rousseeuw, 1987), which yielded a value of 8 for the number of clusters.

c. Model efficiency

To quantify the model predictive skill, a model efficiency statistical score (m_e) was computed for each metric and in each BGC region:

$$m_e = 1 - \frac{\sum_{i=1}^{N}(m_i - o_i)^2}{\sum_{i=1}^{N}(o_i - \bar{o})^2}, \qquad (3)$$

where m_i and o_i are the model and BGC-Argo matched values, respectively, $\bar{o}$ is the BGC-Argo climatology and N is the number of matchups. Assuming that the spatial variations are small in a given BGC-region, $\bar{o}$ represents the temporal average and $\sum_{i=1}^{N}(o_i - \bar{o})^2$ represents the variance due to temporal fluctuations. The model efficiency tests whether the model outperforms the BGC-Argo climatology ($0 < m_e < 1$, Fennel et al., 2022), or stated differently, if the model-data mean square difference is lower than the observation variance, i.e., $\frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2 < \frac{1}{N}\sum_{i=1}^{N}(o_i - \bar{o})^2$. To ensure the robustness of m_e, we verified that the number of matchups for each metric and in each BGC-region was greater than 100, then outliers were removed using Tukey's fences (Tukey, 1977).

4. Results and discussion

a. BGC-regions of the Global Ocean

The K-means clustering algorithm identified 8 distinct BGC-regions (Figure 2). 6 of the 8 BGC-regions correspond to well-defined spatial regions and are, thus, named accordingly, i.e., the Arctic, Equatorial (Equ.), Mediterranean Sea (Med. Sea), OMZs, Subtropical Gyres (Sub. Gyres) and Southern Oceans BGC-regions. The other two BGC-regions, i.e., the Low Nutrients Bloom (Low Nut. Bloom) and the High Nutrients Bloom (High Nut. Bloom) are located in the North Atlantic and North Pacific oceans, as well as in the lower latitudes of the Southern Oceans. These two BGC-regions correspond to ocean basins that experience a phytoplankton bloom in the springtime (Westberry et al., 2016). The main difference between these regions is that in one of them, macronutrients such as nitrate and phosphate are abundant throughout the year due to phytoplankton growth being mainly limited by iron (Williams and Follows, 2011). Finally, it should be noted that outlier grid cells were not removed from the analysis; these outliers are mainly present in grid cells close to the coast. Additionally, grid cells with bathymetry shallower than 1000 m were not included in the clustering analysis as metrics associated with mesopelagic processes cannot be calculated in these shallow grid cells.

The BGC-regions found in our study are overall coherent with the biomes estimated in Fay and McKinley (2014) (hereinafter denoted FM2014). The Arctic and Southern Oceans correspond to the FM2014 ice biome. The Sub. Gyres correspond to the FM2014 subtropical permanently stratified biome. The Equatorial BGC-region represents a larger area than the Equatorial biome in FM2014. The Low Nut. and High Nut. Bloom regions correspond to FM2014 subtropical seasonally stratified and subpolar seasonally stratified biomes, respectively. These two BGC-regions are coherent in the North Pacific and in the Southern Oceans in both studies. They differ, however, in the North Atlantic. In FM2014, the subpolar North Atlantic is divided between the subtropical seasonally stratified and subpolar seasonally stratified biomes, whereas in our study this area is only represented by one BGC-region; the Low Nut. Bloom. Finally, the Med. Sea and OMZs BGC-regions are not represented in FM2014. The main differences between our study and FM2014 are due to differences in the methodology used to identify BGC-regions. In our study, we used 23 input variables to identify BGC-regions, while in FM2014, clustering was based on only one BGC input variable (Chl*a*) and three physical variables (sea surface temperature, MLD, and sea-ice fraction). Our method allows for the identification of specific BGC-regions whose function is mainly characterized by variables other than Chl*a*, such as OMZs. Furthermore, our method includes coastal areas, and identifies the Med. Sea as a BGC-region, which is not included in FM2014 because it is considered a coastal region.

b. Model performance

Figures 3-5 display the model efficiency (*me*) calculated for each assessment metric and BGC region. To enhance clarity, the *me* values are grouped by process, namely carbonate chemistry, biological carbon pump, and oxygen levels. The results are presented as bubble plots (panels b), where the size of the bubble is proportional to the *me* value. A bar plot (panels c) shows the median *me* value for a given assessment metric across all BGC regions, while another bar plot (panels a) shows the median *me* value for a given BGC region across all assessment metrics. Due to the limited number of assessment metrics associated with oxygen levels in most regions (i.e., 2), the mean is used instead of the median. The x and y axes in panels b are arranged in descending order based on the median *me* value across all assessment metrics (as shown in panels a) and the median *me* value across all BGC regions (as shown in panel b), respectively.

i. Carbonate chemistry

The model demonstrates improved performance in predicting certain carbonate chemistry metrics (i.e., Alk_{meso}, DIC_{mixed}, Alk_{mixed}, DIC_{meso}, and pH_{meso}) compared to the BGC-Argo climatology, as indicated by median *me* values significantly greater than 0 (Figs. 3b and c). However, the model's ability to reproduce instantaneous variability in pH_{mixed} is more limited, with a *me* value close to 0, indicating no improvement over a simple average of observed values. Furthermore, the model underperforms the BGC-Argo climatology for pCO_{2mixed} across all regions. Despite these limitations, the model provides an overall better estimate of carbonate chemistry dynamics in all BGC regions compared to the BGC-Argo climatology, as evidenced by Figure 3a.

ii. Biological carbon pump

The efficiency of the model in estimating the biological carbon pump metrics varies across both metrics and regions (Fig. 4a-c). The model outperforms the BGC-Argo climatology in estimating PO_4 and NO_3 in the mesopelagic and mixed layer, as well as Si_{meso} and H_{nit}.

However, the model's ability to predict Si decreases significantly as one moves from the mesopelagic to the mixed layer. Additionally, the metrics associated with the first trophic level, such as Chl_{mixed}, H_{DCM}, Chl_{DCM}, POC_{mixed}, and POC_{meso}, are systematically outperformed by the BGC-Argo climatology, with median *me* values less than 0 in nearly all BGC regions (Figure 4b). Regional analysis of the median *me* values (Figure 4a) shows that the model performs better than the observational mean (median *me* values greater than 0) in only a few regions (i.e., the High Nut. Bloom, the Low Nut. Bloom, the Med. Sea, and the OMZs) indicating that the model performs relatively well in these regions, but may not be as accurate in the other regions.

iii. Oxygen levels

The model provides better estimates of mixed and mesopelagic O_2 concentrations in most BGC regions compared to the BGC-Argo climatology, as evidenced by consistently positive *me* values in Figure 5b. However, the BGC-Argo climatology provides a better representation of the magnitude of O_{2min} compared to the model, while the model performs better than the climatology in predicting H_{O2min}, but only in the OMZs BGC-region. These results suggest that while the model performs well in estimating mixed and mesopelagic O_2 concentrations in most BGC regions, it doesn't accurately capture the depth and magnitude of OMZs.

iv. Discussion

The model outperforms the BGC-Argo climatology for DIC, Alk, NO_3, PO_4, in the mesopelagic layer and mixed layers and Si in the mesopelagic layer. We attribute this good performance to the effective application of climatological damping. As described in the Appendix, the climatological damping mitigates the effects of physical data assimilation in the offline coupled hydrodynamic-biogeochemical system, which can lead to unrealistic drift of various biogeochemical variables. Specifically, we used the World Ocean Atlas 2013 (Garcia et al., 2013, 2014) for NO_3, PO_4, O_2, and Si, and the Global Ocean Data Analysis Project version 2 (GLODAPv2) climatology (Key et al., 2015) for DIC and Alk. However, our analysis revealed that the model's performance in estimating Si in the mixed layer is significantly degraded comparing to the mesopelagic layer, indicating the presence of additional factors affecting the model's ability to accurately estimate this variable. Further investigation is required to identify these factors and improve the model's performance in estimating Si in the mixed layer.

For the three Chla-related metrics, the model performs worse than the BGC-Argo climatology. This is unexpected, as the model incorporates a reduced-order Kalman filter Lellouche et al., 2013) that assimilates daily L4 remotely sensed surface Chla, providing a mixed-layer correction to the modeled Chla (see Appendix). We verified that the assimilation of satellite Chla improves the model's ability to predict Chla, as the model-BGC-Argo data misfit is lower compared to a simulation without assimilation (not shown). Furthermore, the model-satellite misfit was also found to be lower than the variability of the satellite data (European Union-Copernicus Marine Service, 2019). These results suggest that discrepancies between the assimilated satellite Chla product and the BGC-Argo data may be responsible for the observed model-BGC-Argo data misfit. Therefore, we suggest that future studies investigate the consistency between ocean colour products and BGC-Argo Chl*a* products on a global scale, as these two products are expected to be assimilated together in future operational BGC systems (Ford, 2021).

Overall, the model also performs worse or no better than the BGC-Argo climatology in predicting POC concentrations, the magnitude and depth of OMZs, pH_{mixed} and $pCO_{2\ mixed}$. The poor performance of PISCES-based simulations relative to BGC-Argo POC observations has been extensively studied in Gali et al. (2022). They pointed out that the large model-data misfit could be the result of an imperfect BGC-Argo POC-b_{bp} conversion factor, unsuitable model parameters associated with POC dynamics and missing processes in the model structure. Similarly, the poor model skill in capturing the OMZs dynamics has also already been documented in several studies (Busecke et al., 2022; Schmidt et al., 2021; Cabré et al., 2015). All these studies suggested that improving the ocean circulation in physical models may be the most important factor to improve the accuracy of OMZs model predictions. Finally, the negative model efficiencies for pH_{mixed} and $pCO_{2\ mixed}$ can be attributed to the fact that these variables are driven by DIC, Alk, temperature, and salinity. Therefore, even small uncertainties in the model estimates of DIC, Alk (as shown in Figure 3b), temperature, and salinity (Lellouche et al., 2018) can result in poor model performance in capturing the variability of pH and pCO_2. This highlights the importance of accurately modelling these four variables to improve model estimates of pH and pCO_2.

c. Recommendation for the design of the BGC-Argo observing system

Observing System Simulation Experiments (OSSE) have been the primary tool to inform about the design of the BGC-Argo observing system (Ford, 2021; Biogeochemical-Argo Planning Group, 2016). OSSEs typically comprises a realistic "nature run", which represents "the truth" from which synthetic observations are sampled. The synthetic observations represent the observing system to be designed. To test its impact on improving model's predictive skill, the synthetic observations are then assimilated in an "assimilative run". The accuracy of the "assimilative run" is then evaluated against the "nature run". Here, we use the real BGC-Argo observations to inform about the design of the BGC-Argo network. More specifically, our aim is to inform about the regions where the model errors are greater than the variability of the BGC-Argo data, and consequently where BGC-Argo observations should be enhanced to improve the model accuracy through BGC-Argo data assimilation or process-oriented assessment studies.

For a given BGC-region, we compute a single multivariate score corresponding to the median of the 23 *me* associated with each assessment metric (Fig. 6). This is consistent with the fact that the BGC-Argo floats, that are now deployed, observe the 5 variables used to derive the assessment metrics, i.e., O_2, Chl_a, NO3, b_{bp} and pH. In the Arctic and in the Southern Ocean BGC-regions (typically North of 60°N and South of 60°S), the median *me* is barely greater than 0, suggesting that in these regions, the model performs no better than a simple mean of the observed values. In these two regions, the model is not well constrained by the assimilation of remotely sensed Chl*a* because satellite observations of ocean colour are not possible for most of the year due to ubiquitous cloud cover. In addition, the lack of in situ observations makes the climatological forcing less efficient in these regions. Together, these factors are likely to lead to poorer model performance compared to other regions. Consequently, we strongly recommend enhancing the Arctic region where BGC-Argo observations are scarce (Fig.1), and where the winter-spring months are particularly under-sampled (not shown). We also recommend maintaining the already-high-density of BGC-Argo observations in the Southern Ocean. These observations are critical to better constrain the model in these two regions where the constraint of models by assimilation of satellite observations is not possible for most of the year.

5. Conclusion

In this study, we propose a method based on the global data set of BGC-Argo observations, a K-means clustering algorithm and 23 assessment metrics to simplify model-data comparison and inform on Copernicus Marine Service forecasting system predictive skill and the design of the BGC-Argo observing system. The K-means algorithm identified 8 BGC-regions in the model simulation that are consistent with Fay and McKinley (2014) study. Within each BGC-region and for each assessment metric, we compute a model efficiency statistical score that quantifies whether the model outperforms the BGC-Argo climatology by comparing the model-BGC-Argo data mean square difference with the observation variance.

Overall, the model surpasses the BGC-Argo climatology in predicting pH, DIC, Alk, O_2, NO_3 and PO4 in the mesopelagic and the mixed layers, as well as Si in the mesopelagic layer. For the other metrics, whose model predictions are outperformed by the BGC-Argo climatology, we provide suggestions to reduce the model-data misfit and thus to increase the model efficiency. Regarding the estimation of Si in the mixed layer, we suggest the presence of additional factors that may affect the model's ability to accurately estimate this variable. Further investigation is necessary to identify these factors and improve the model's performance in this regard. For Chl*a*-related metrics, we recommend checking the consistency between ocean colour products and BGC-Argo Chl*a* products at the global scale as it may explain part of the misfit between the model, that assimilates satellite Chl*a*, and BGC-Argo observations. The discrepancies between modelled and observed POC and OMZs have been already investigated in previous studies. It has been suggested that improving the BGC-Argo POC-b_{bp} conversion factor, tuning the model parameters and implementing missing processes in the model structure could decrease the model-data inconsistencies associated with POC dynamics. Similarly, improving the ocean circulation in physical models should improve the accuracy of OMZ model predictions. Finally, pH_{mixed} and $pCO_{2\ mixed}$ should be better modelled if the uncertainties associated with DIC, Alk, temperature and salinity in the mixed layer are reduced.

The proposed method can also be used to optimize the design of the BGC-Argo network. In particular, the regions where BGC-Argo observations should be enhanced to reduce the model-data misfit through the assimilation of BGC-Argo data or process-oriented assessment studies. We strongly recommend enhancing the observation density in the Arctic region and maintaining the already high density of observations in the Southern Oceans. These are two regions where the model error is barely less than the variability of BGC-Argo observations, and where it is not possible to use satellite observations to constrain the models through assimilation most of the year.

Tables

Table 1. Data mode and QC flags of the BGC-Argo observations used in this study. In the Argo data-system, the data are available in three data modes: "Real-Time", "Adjusted" and "Delayed". See section 2a for a brief description of each data mode. The flags "3" and "4" refer to "potentially bad data" and "bad data", respectively. See also Bittig et al. (2019), for a more detailed description of Argo data modes and flags.

Parameter	Data mode	Data mode of associated pressure, temperature and salinity profiles	QC flags
Chl*a*	Adjusted and Delayed	Real time, Adjusted and Delayed	• Real time (P,T,S): All flags except 4 • Adjusted or Delayed: All flags except 3 and 4
O_2	Delayed	Delayed	• All flags except 3 and 4
NO_3	Adjusted and Delayed	Real time, Adjusted and Delayed	• Real time (P,T,S): All flags except 4 • Adjusted or Delayed: All flags except 3 and 4
pH	Adjusted and Delayed	Real time, Adjusted and Delayed	• Real time (P,T,S): All flags except 4 • Adjusted or Delayed: All flags except 3 and 4
b_{bp}	Adjusted and Delayed	Real time, Adjusted and Delayed	• Real time (P,T,S): All flags except 4 • Adjusted or Delayed (P,T,S): All flags except 3 and 4

- Adjusted or Delayed (b_{bp}): All flags except 4

Table 2. Assessment metrics used to assess the model simulation with BGC-Argo data. For each metric, the level of assessment, as described in Hipsey et al. (2020) is also indicated.

Process	Metric	Definition	units	Assessment level
Carbonate chemistry	$pCO_{2\ mixed}$	Depth-averaged pCO_2 in the mixed layer	µatm	State variable
	DIC_{mixed}	Depth-averaged DIC in the mixed layer	µmol kg^{-1}	State variable
	Alk_{mixed}	Depth-averaged Alk in the mixed layer	µmol kg^{-1}	State variable
	DIC_{meso}	Depth-averaged DIC in the mesopelagic layer	µmol kg^{-1}	State variable
	Alk_{meso}	Depth-averaged Alk in the mesopelagic layer	µmol kg^{-1}	State variable
	pH_{mixed}	Depth-averaged pH in the mixed layer	total	State variable
	pH_{meso}	Depth-averaged pH in the mesopelagic layer	total	State variable
Biological carbon pump	Chl_{mixed}	Depth-averaged Chl*a* in the mixed layer	mg m^{-3}	State variable
	$NO_{3\ mixed}$	Depth-averaged NO_3 in the mixed layer	µmol kg^{-1}	State variable
	$PO_{4\ mixed}$	Depth-averaged PO_4 in the mixed layer	µmol kg^{-1}	State variable
	Si_{mixed}	Depth-averaged Si in the mixed layer	µmol kg^{-1}	State variable

	$NO_{3\ meso}$	Depth-averaged NO_3 in the mesopelagic layer	µmol kg^{-1}	State variable
	$PO_{4\ meso}$	Depth-averaged PO_4 in the mesopelagic layer	µmol kg^{-1}	State variable
	Si_{meso}	Depth-averaged Si in the mesopelagic layer	µmol kg^{-1}	State variable
	POC_{mixed}	Depth-averaged POC in the mixed layer	mg m^{-3}	State variable
	POC_{meso}	Depth-averaged POC in the mesopelagic layer	mg m^{-3}	State variable
	Chl_{DCM}	Magnitude of DCM	mg m^{-3}	Emergent property
	H_{DCM}	Depth of DCM	m	Emergent property
	H_{nit}	Depth of nitracline	m	Emergent property
Oxygen levels	$O_{2\ mixed}$	Depth-averaged O_2 in the mixed layer	µmol kg^{-1}	State variable
	$O_{2\ meso}$	Depth-averaged O_2 in the mesopelagic layer	µmol kg^{-1}	State variable
	O_{2min}	Value of O_2 minimum	µmol kg^{-1}	Emergent property
	H_{O2min}	Depth of O_2 minimum	m	Emergent property

Figures

Figure 1. Spatial and temporal coverage of BGC-Argo quality- controlled pH, NO_3, Chl*a*, O_2 and b_{bp} profiles. **(a)** Number of quality-controlled profiles for the entire period per 4°x4° bin. **(b)** Number of quality-controlled profiles per year. Note that this study only uses data from 2009 to 2020 to evaluate model performance.

Figure 2. Spatial distribution of the 8 BGC-regions obtained with a K-means clustering method applied to a dataset of modelled climatological monthly time series of the 23 assessment metrics.

Figure 3. Bubble plot of model efficiency statistical score (m_e) as a function of BGC-regions [Arctic, Equatorial (Equ.), High Nutrients Bloom (High Nut. Bloom), Low Nutrients Bloom (Low Nut. Bloom), Mediterranean Sea (Med. Sea), Oxygen Minimum Zones (OMZs), Southern Oceans, Subtropical Gyres (Sub. Gyres)] and assessment metrics associated with the carbonate chemistry [depth-averaged pCO_2 in the mixed layer ($pCO_{2\ mixed}$), depth-averaged DIC in the mixed layer (DIC_{mixed}), depth-averaged Alk in the mixed layer (Alk_{mixed}), depth-averaged DIC in the mesopelagic layer (DIC_{meso}), depth-averaged Alk in the mesopelagic layer (Alk_{meso}), depth-averaged pH in the mixed layer (pH_{mixed}), and depth-averaged pH in the mesopelagic layer (pH_{meso})] (**b**). The size of a bubble is proportional to the value of m_e. For a given assessment metric, the median values of m_e over all BGC regions are represented as a bar plot (**c**). Similarly, for a given BGC region, the median values of m_e across all assessment metrics are represented as a bar plot (**a**). In (**b**), The x and y axes are arranged in descending order of the median value of m_e over all assessment metrics (panels a) and the median value of m_e over all BGC regions, respectively. The blue and orange colours correspond to a positive and negative m_e .

Figure 4. Same as Figure 3 but for assessment metrics associated with the biological carbon pump [depth-averaged Chl*a* in the mixed layer (Chl_{mixed}), depth-averaged NO_3 in the mixed layer ($NO_{3\ mixed}$), depth-averaged PO_4 in the mixed layer ($PO_{4\ mixed}$), depth-averaged Si in the mixed layer (Si_{mixed}), depth-averaged NO_3 in the mesopelagic layer ($NO_{3\ meso}$), depth-averaged PO_4 in the mesopelagic layer ($PO_{4\ meso}$), depth-averaged Si in the mesopelagic layer (Si_{meso}), depth-averaged POC in the mixed layer (POC_{mixed}), depth-averaged POC in the mesopelagic layer (POC_{meso}), magnitude of DCM (Chl_{DCM}), depth of DCM (H_{DCM}), and depth of nitracline (H_{nit})].

Figure 5. Same as Figure 3 but for assessment metrics associated with the oxygen levels [depth-averaged O_2 in the mixed layer ($O_{2\ mixed}$), depth-averaged O_2 in the mesopelagic layer ($O_{2\ meso}$), value of O_2 minimum (O_{2min}), and depth of O_2 minimum (H_{O2min})]. Note that in **(a)**, the bar plot represents the mean value of m_e over all assessment metrics.

Figure 6. Median of the 23 m_e associated with each assessment metric by BGC-region.

Appendix

A.1 The CMEMS global hydrodynamic-biogeochemical model

The model used in this study features the offline coupled NEMO–PISCES model, with a 1/4° horizontal resolution 50 vertical levels (with 22 levels in the upper 100 m, the vertical resolution is 1m near the surface and decreases to 450m resolution near the bottom) and daily temporal resolution, covering the period from 2009 to 2017.

The biogeochemical model PISCES v2 (Aumont et al., 2015) is a model of intermediate complexity designed for global ocean applications, and is part of NEMO modelling platform. It features 24 prognostic variables and includes five nutrients that limit phytoplankton growth (nitrate, ammonium, phosphate, silicate and iron) and four living compartments: two phytoplankton size classes (nanophytoplankton and diatoms, resp. small and large) and two zooplankton size classes (microzooplankton and mesozooplankton, resp. small and large); the bacterial pool is not explicitly modelled. PISCES distinguishes three non-living detrital pools for organic carbon, particles of calcium carbonate and biogenic silicate. Additionally, the model simulates the carbonate system and dissolved oxygen. PISCES has been successfully used in a variety of biogeochemical studies, both at regional and global scale (Bopp et al., 2005; Gehlen et al., 2006, 2007; Gutknecht et al., 2019; Lefèvre et al., 2019; Schneider et al., 2008; Séférian et al., 2013; Steinacher et al., 2010; Tagliabue et al., 2010).

The dynamical component is the latest Mercator Ocean global 1/12° high-resolution ocean model system, extensively described and validated in Lellouche et al. (2013, 2018). This system provides daily and 1/4°-coarsened fields of horizontal and vertical current velocities, vertical eddy diffusivity, mixed layer depth, sea ice fraction, potential temperature, salinity, sea surface height, surface wind speed, freshwater fluxes and net surface solar shortwave irradiance that drive the transport of biogeochemical tracers. This system also features a reduced-order Kalman filter based on the Singular Evolutive Extended Kalman filter (SEEK) formulation introduced by Pham et al. (1998), that assimilates, on a 7-day assimilation cycle, along-track altimeter data, satellite Sea Surface Temperature and Sea-Ice Concentration from

OSTIA, and *in situ* temperature and salinity vertical profiles from the CORA 4.2 in situ database.

In addition, the biogeochemical component of the coupled system also embeds a reduced order Kalman filter (similar to the above mentioned) that operationally assimilates daily L4 remotely sensed surface chlorophyll (European Union-Copernicus Marine Service, 2022). Thanks to a multivariate formulation of model error covariances, the system is able to provide a 3D correction to the nanophytoplankton, diatoms and nitrates model concentrations, from the surface chlorophyll data provided by satellite observations.

In parallel, a climatological-damping is applied to nitrate, phosphate, oxygen, silicate - with World Ocean Atlas 2013 - to dissolved inorganic carbon and alkalinity – with GLODAPv2 climatology (Key et al., 2015) - and to dissolved organic carbon and iron - with a 4000-year PISCES climatological run. This relaxation is set to mitigate the impact of the physical data assimilation in the offline coupled hydrodynamic-biogeochemical system, leading significant rises of nutrients in the Equatorial Belt area, and resulting in an unrealistic drift of various biogeochemical variables e.g. chlorophyll, nitrate, phosphate (Fennel et al., 2019; Park et al., 2018). The time-scale associated with this climatological damping is set to 1 year and allows a smooth constraint that has been shown to be efficient to reduce the model drift.

Data availability. The BGC model data can be downloaded from the Copernicus Marine Environmental Monitoring Service (https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=GLOBAL_ANALYSIS_FORECAST_BIO_001_028). The BGC-Argo data were downloaded from the Argo Global Data Assembly Centre in France (ftp://ftp.ifremer.fr/argo/).

Authors Contribution: AM, GC, FD, SS and VT originated the study. AM, HC, FD, RS and VT designated the study. AM and RS process the BGC-Argo floats data. AM analysed the data. AM wrote the first draft of the manuscript. HC, GC, FD, EG, PL, CP, SS,RS,VT and AT contributed to the subsequent drafts. All authors read and approved the final draft.

Competing Interests: The authors declare no competing financial interests.

Materials and correspondence: Correspondence and request for material should be addressed to mignot@mercator-ocean.fr

Acknowledgements: This study has been conducted using the Copernicus Marine Service products. The BGC-Argo data were collected and made freely available by the International Argo program and the national programs that contribute to it (https://www.argo.jcommops.org). The Argo program is part of the Global Ocean Observing System. Part of this work was performed within the framework of the BIOOPTIMOD and MASSIMILI CMEMS Service Evolution Projects. This paper represents a contribution to the following research projects: NAOS (funded by the Agence Nationale de la Recherche in the framework of the French "Equipement d'avenir" program, grant ANR J11R107-F), remOcean (funded by the European Research Council, grant 246777), and the French Bio-Argo program (BGC-Argo France; funded by CNES-TOSCA, LEFE-GMMC).

References

Allen, J. I., Somerfield, P. J., and Gilbert, F. J.: Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models, J. Mar. Syst., 64, 3–14, https://doi.org/10.1016/j.jmarsys.2006.02.010, 2007.

Álvarez, E., Lazzari, P., and Cossarini, G.: Phytoplankton diversity emerging from chromatic adaptation and competition for light, Prog. Oceanogr., 204, 102789, https://doi.org/10.1016/j.pocean.2022.102789, 2022.

CMEMS: https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=GLOBAL_ANALYSIS_FORECAST_BIO_001_028, last access: 29 October 2020.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geosci. Model Dev., 8, 2465–2513, https://doi.org/10.5194/gmd-8-2465-2015, 2015.

Barbieux, M., Uitz, J., Gentili, B., Pasqueron de Fommervault, O., Mignot, A., Poteau, A., Schmechtig, C., Taillandier, V., Leymarie, E., Penkerc'h, C., D'Ortenzio, F., Claustre, H., and Bricaud, A.: Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database, Biogeosciences, 16, 1321–1342, https://doi.org/10.5194/bg-16-1321-2019, 2019.

Biogeochemical-Argo Planning Group: The scientific rationale, design and implementation plan for a Biogeochemical-Argo float array, https://doi.org/10.13155/46601, 2016.

Bittig, H. C., Steinhoff, T., Claustre, H., Fiedler, B., Williams, N. L., Sauzède, R., Körtzinger, A., and Gattuso, J.-P.: An alternative to static climatologies: robust estimation of open ocean CO2 variables and nutrient concentrations from T, S, and O2 data using Bayesian neural networks, Front. Mar. Sci., 5, 328, 2018.

Bittig, H. C., Maurer, T. L., Plant, J. N., Wong, A. P., Schmechtig, C., Claustre, H., Trull, T. W., Udaya Bhaskar, T. V. S., Boss, E., and Dall'Olmo, G.: A BGC-Argo guide: Planning, deployment, data handling and usage, Front. Mar. Sci., 6, 502, 2019.

Bock, N., Cornec, M., Claustre, H., and Duhamel, S.: Biogeographical Classification of the Global Ocean From BGC-Argo Floats, Glob. Biogeochem. Cycles, 36, https://doi.org/10.1029/2021GB007233, 2022.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of essential climate variables in support of climate research, applications, and policy, Bull. Am. Meteorol. Soc., 95, 1431–1443, 2014.

Bopp, L., Aumont, O., Cadule, P., Alvain, S., and Gehlen, M.: Response of diatoms distribution to global warming and potential implications: A global model study, Geophys. Res. Lett., 32, https://doi.org/10.1029/2005GL023653, 2005.

Boyer, T. P., Antonov, J. I., Baranova, O. K., Garcia, H. E., Johnson, D. R., Mishonov, A. V., O'Brien, T. D., Seidov, D., Smolyar, I., and Zweng, M. M.: World ocean database 2013, 2013.

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., Garçon, V., Gilbert, D., Gutiérrez, D., Isensee, K., Jacinto, G. S., Limburg, K. E., Montes, I., Naqvi, S. W. A., Pitcher, G. C., Rabalais, N. N., Roman, M. R., Rose, K. A., Seibel, B. A., Telszewski, M., Yasuhara, M., and Zhang, J.: Declining oxygen in the global ocean and coastal waters, Science, 359, https://doi.org/10.1126/science.aam7240, 2018.

Briggs, N., Perry, M. J., Cetinić, I., Lee, C., D'Asaro, E., Gray, A. M., and Rehm, E.: High-resolution observations of aggregate flux during a sub-polar North Atlantic spring bloom, Deep Sea Res. Part Oceanogr. Res. Pap., 58, 1031–1039, https://doi.org/10.1016/j.dsr.2011.07.007, 2011.

Busecke, J. J. M., Resplandy, L., Ditkovsky, S. J., and John, J. G.: Diverging Fates of the Pacific Ocean Oxygen Minimum Zone and Its Core in a Warming World, AGU Adv., 3, https://doi.org/10.1029/2021AV000470, 2022.

Cabré, A., Marinov, I., Bernardello, R., and Bianchi, D.: Oxygen minimum zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends, Biogeosciences, 12, 5429–5454, https://doi.org/10.5194/bg-12-5429-2015, 2015.

Capuzzo, E., Lynam, C. P., Barry, J., Stephens, D., Forster, R. M., Greenwood, N., McQuatters-Gollop, A., Silva, T., van Leeuwen, S. M., and Engelhard, G. H.: A decline in primary production in the North Sea over 25 years, associated with reductions in zooplankton abundance and fish stock recruitment, Glob. Change Biol., 24, e352–e364, https://doi.org/10.1111/gcb.13916, 2018.

Cermeno, P., Dutkiewicz, S., Harris, R. P., Follows, M., Schofield, O., and Falkowski, P. G.: The role of nutricline depth in regulating the ocean carbon cycle, Proc. Natl. Acad. Sci., 105, 20344–20349, https://doi.org/10.1073/pnas.0811302106, 2008.

Claustre, H., Johnson, K. S., and Takeshita, Y.: Observing the Global Ocean with Biogeochemical-Argo, Annu. Rev. Mar. Sci., 12, annurev-marine-010419-010956, https://doi.org/10.1146/annurev-marine-010419-010956, 2020.

Crowder, L. B., Hazen, E. L., Avissar, N., Bjorkland, R., Latanich, C., and Ogburn, M. B.: The Impacts of Fisheries on Marine Ecosystems and the Transition to Ecosystem-Based Management, Annu. Rev. Ecol. Evol. Syst., 39, 259–278, https://doi.org/10.1146/annurev.ecolsys.39.110707.173406, 2008.

Cullen, J. J.: Subsurface Chlorophyll Maximum Layers: Enduring Enigma or Mystery Solved?, Annu. Rev. Mar. Sci., 7, 207–239, https://doi.org/10.1146/annurev-marine-010213-135111, 2015.

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., Glover, D. M., and Takahashi, T.: Skill metrics for confronting global upper ocean ecosystem-biogeochemistry models against field and remote sensing data, J. Mar. Syst., 76, 95–112, https://doi.org/10.1016/j.jmarsys.2008.05.015, 2009.

D'Ortenzio, F. and d'Alcala, M. R.: On the trophic regimes of the Mediterranean Sea: a satellite analysis, Biogeosciences, 6, 139–148, 2009.

D'Ortenzio, F., Taillandier, V., Claustre, H., Prieur, L. M., Leymarie, E., Mignot, A., Poteau, A., Penkerc'h, C., and Schmechtig, C. M.: Biogeochemical Argo: The Test Case of the NAOS Mediterranean Array, Front. Mar. Sci., 7, 120, https://doi.org/10.3389/fmars.2020.00120, 2020.

Dutkiewicz, S., Hickman, A. E., Jahn, O., Gregg, W. W., Mouw, C. B., and Follows, M. J.: Capturing optically important constituents and properties in a marine biogeochemical and ecosystem model, Biogeosciences, 12, 4447–4481, https://doi.org/10.5194/bg-12-4447-2015, 2015.

Eriksen, M., Lebreton, L. C. M., Carson, H. S., Thiel, M., Moore, C. J., Borerro, J. C., Galgani, F., Ryan, P. G., and Reisser, J.: Plastic Pollution in the World's Oceans: More than 5 Trillion Plastic Pieces Weighing over 250,000 Tons Afloat at Sea, PLoS ONE, 9, e111913, https://doi.org/10.1371/journal.pone.0111913, 2014.

European Union-Copernicus Marine Service: Global Ocean- In-Situ Near-Real-Time Observations, https://doi.org/10.48670/MOI-00036, 2015.

European Union-Copernicus Marine Service: Global Ocean Biogeochemistry Analysis and Forecast, https://doi.org/10.48670/MOI-00015, 2019.

European Union-Copernicus Marine Service: Global Ocean 3D Chlorophyll-a concentration, Particulate Backscattering coefficient and Particulate Organic Carbon, https://doi.org/10.48670/MOI-00046, 2020.

European Union-Copernicus Marine Service: Global Ocean Colour (Copernicus-GlobColour), Bio-Geo-Chemical, L4 (monthly and interpolated) from Satellite Observations (Near Real Time), https://doi.org/10.48670/MOI-00279, 2022.

Fay, A. R. and McKinley, G. A.: Global open-ocean biomes: mean and temporal variability, Earth Syst. Sci. Data, 6, 273–284, https://doi.org/10.5194/essd-6-273-2014, 2014.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A. M., Gregoire, M., Jones, E., Kim, H.-C., Lamouroux, J., Murtugudde, R., Perruche, C., and the GODAE OceanView Marine Ecosystem Analysis and Prediction Task Team: Advancing Marine Biogeochemical and Ecosystem Reanalyses and Forecasts as Tools for Monitoring and Managing Ecosystem Health, Front. Mar. Sci., 6, 89, https://doi.org/10.3389/fmars.2019.00089, 2019.

Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., and Yu, L.: Ocean biogeochemical modelling, Nat. Rev. Methods Primer, 2, 1–21, https://doi.org/10.1038/s43586-022-00154-2, 2022.

Ford, D.: Assimilating synthetic Biogeochemical-Argo and ocean colour observations into a global ocean model to inform observing system design, Biogeosciences, 18, 509–534, https://doi.org/10.5194/bg-18-509-2021, 2021.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., Arneth, A., Arora, V. K., Bates, N. R., Becker, M., Bellouin, N., Bittig, H. C., Bopp, L., Chevallier, F., Chini, L. P., Cronin, M., Evans, W., Falk, S., Feely, R. A., Gasser, T., Gehlen, M., Gkritzalis, T., Gloege, L., Grassi, G., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jain, A. K., Jersild, A., Kadono, K., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lindsay, K., Liu, J., Liu, Z., Marland, G., Mayot, N., McGrath, M. J., Metzl, N., Monacci, N. M., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K., Ono, T., Palmer, P. I., Pan, N., Pierrot, D., Pocock, K., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Rodriguez, C., Rosan, T. M., Schwinger, J., Séférian, R., Shutler, J. D., Skjelvan, I., Steinhoff, T., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tanhua, T., Tans, P. P., Tian, X., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., Walker, A. P., Wanninkhof, R., Whitehead, C., Willstrand Wranne, A., et al.: Global Carbon Budget 2022, Earth Syst. Sci. Data, 14, 4811–4900, https://doi.org/10.5194/essd-14-4811-2022, 2022.

Galí, M., Falls, M., Claustre, H., Aumont, O., and Bernardello, R.: Bridging the gaps between particulate backscattering measurements and modeled particulate organic carbon in the ocean, Biogeosciences, 19, 1245–1275, https://doi.org/10.5194/bg-19-1245-2022, 2022.

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

Garcia, H. E., Boyer, T. P., Locarnini, R. A., Antonov, J. I., Mishonov, A. V., Baranova, O. K., Zweng, M. M., Reagan, J. R., Johnson, D. R., and Levitus, S.: World ocean atlas 2013. Volume 3, Dissolved oxygen, apparent oxygen utilization, and oxygen saturation, 2014.

Gasparin, F., Cravatte, S., Greiner, E., Perruche, C., Hamon, M., Van Gennip, S., and Lellouche, J.-M.: Excessive productivity and heat content in tropical Pacific analyses: Disentangling the effects of in situ and altimetry assimilation, Ocean Model., 160, 101768, https://doi.org/10.1016/j.ocemod.2021.101768, 2021.

Gehlen, M., Bopp, L., Emprin, N., Aumont, O., Heinze, C., and Ragueneau, O.: Reconciling surface ocean productivity, export fluxes and sediment composition in a global biogeochemical ocean model, Biogeosciences, 3, 521–537, https://doi.org/10.5194/bg-3-521-2006, 2006.

Gehlen, M., Gangstø, R., Schneider, B., Bopp, L., Aumont, O., and Ethe, C.: The fate of pelagic $CaCO_3$ production in a high CO_2 ocean: a model study, Biogeosciences, 4, 505–519, https://doi.org/10.5194/bg-4-505-2007, 2007.

Gutknecht, E., Reffray, G., Mignot, A., Dabrowski, T., and Sotillo, M. G.: Modelling the marine ecosystem of Iberia-Biscay-Ireland (IBI) European waters for CMEMS operational applications, Ocean Sci., 15, 1489–1516, https://doi.org/10.5194/os-15-1489-2019, 2019.

Hartigan, J. A. and Wong, M. A.: Algorithm AS 136: A K-Means Clustering Algorithm, Appl. Stat., 28, 100, https://doi.org/10.2307/2346830, 1979.

Hipsey, M. R., Gal, G., Arhonditsis, G. B., Carey, C. C., Elliott, J. A., Frassl, M. A., Janse, J. H., de Mora, L., and Robson, B. J.: A system of metrics for the assessment and improvement of aquatic ecosystem models, Environ. Model. Softw., 128, 104697, https://doi.org/10.1016/j.envsoft.2020.104697, 2020.

Iida, Y., Takatani, Y., Kojima, A., and Ishii, M.: Global trends of ocean CO 2 sink and ocean acidification: an observation-based reconstruction of surface ocean inorganic carbon variables, J. Oceanogr., 1–36, 2020.

Johnson, Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., Swift, D. D., Williams, N. L., Boss, E., Haëntjens, N., Talley, L. D., and Sarmiento, J. L.: Biogeochemical sensor performance in the SOCCOM profiling float array: SOCCOM BIOGEOCHEMICAL SENSOR PERFORMANCE, J. Geophys. Res. Oceans, 122, 6416–6436, https://doi.org/10.1002/2017JC012838, 2017.

Johnson, Plant, J. N., and Maurer, T. L.: Processing BGC-Argo pH data at the DAC level, 2018a.

Johnson, Pasqueron De Fommervault, O., Serra, R., D'Ortenzio, F., Schmechtig, C., Claustre, H., and Poteau, A.: Processing Bio-Argo nitrate concentration at the DAC Level, 2018b.

Key, R. M., Olsen, A., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., and Hoppema, M.: Global Ocean Data Analysis Project, Version 2 (GLODAPv2), Carbon Dioxide Information Analysis Center, Oak Ridge Nat Lab, 2015.

Kheireddine, M., Mayot, N., Ouhssain, M., and Jones, B. H.: Regionalization of the Red Sea Based on Phytoplankton Phenology: A Satellite Analysis, J. Geophys. Res. Oceans, 126, https://doi.org/10.1029/2021JC017486, 2021.

Lacour, L., Claustre, H., Prieur, L., and D'Ortenzio, F.: Phytoplankton biomass cycles in the North Atlantic subpolar gyre: A similar mechanism for two different blooms in the Labrador Sea: THE LABRADOR SEA BLOOMS, Geophys. Res. Lett., 42, 5403–5410, https://doi.org/10.1002/2015GL064540, 2015.

Lavigne, H., D'Ortenzio, F., Migon, C., Claustre, H., Testor, P., d'Alcalà, M. R., Lavezza, R., Houpert, L., and Prieur, L.: Enhancing the comprehension of mixed layer depth control on the Mediterranean phytoplankton phenology: Mediterranean Phytoplankton Phenology, J. Geophys. Res. Oceans, 118, 3416–3430, https://doi.org/10.1002/jgrc.20251, 2013.

Lazzari, Solidoro, C., Ibello, V., Salon, S., Teruzzi, A., Béranger, K., Colella, S., and Crise, A.: Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: a modelling approach, Biogeosciences, 9, 217–233, https://doi.org/10.5194/bg-9-217-2012, 2012.

Lazzari, Solidoro, C., Salon, S., and Bolzon, G.: Spatial variability of phosphate and nitrate in the Mediterranean Sea: A modeling approach, Deep Sea Res. Part Oceanogr. Res. Pap., 108, 39–52, https://doi.org/10.1016/j.dsr.2015.12.006, 2016.

Lefèvre, N., Veleda, D., Tyaquiçã, P., Perruche, C., Diverrès, D., and Ibánhez, J. S. P.: Basin-Scale Estimate of the Sea-Air CO_2 Flux During the 2010 Warm Event in the Tropical North Atlantic, J. Geophys. Res. Biogeosciences, 124, 973–986, https://doi.org/10.1029/2018JG004840, 2019.

Lellouche, Greiner, E., Le Galloudec, O., Garric, G., Regnier, C., Drevillon, M., Benkiran, M., Testut, C.-E., Bourdalle-Badie, R., Gasparin, F., Hernandez, O., Levier, B., Drillet, Y., Remy, E., and Le Traon, P.-Y.: Recent updates to the Copernicus Marine Service global ocean monitoring and forecasting real-time 1⁄12° high-resolution system, Ocean Sci., 14, 1093–1126, https://doi.org/10.5194/os-14-1093-2018, 2018.

Lellouche, J.-M., Le Galloudec, O., Drévillon, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.-E., Bricaud, C., Bourdallé-Badie, R., Tranchant, B., Benkiran, M., Drillet, Y., Daudin, A., and De Nicola, C.: Evaluation of global monitoring and forecasting systems at Mercator Océan, Ocean Sci., 9, 57–81, https://doi.org/10.5194/os-9-57-2013, 2013.

Letelier, R. M., Karl, D. M., Abbott, M. R., and Bidigare, R. R.: Light driven seasonal patterns of chlorophyll and nitrate in the lower euphotic zone of the North Pacific Subtropical Gyre, Limnol. Oceanogr., 49, 508–519, 2004.

Lynch, D. R., McGillicuddy, D. J., and Werner, F. E.: Skill assessment for coupled biological/physical models of marine systems, J. Mar. Syst., 1, 1–3, 2009.

Macías, D., Stips, A., and Garcia-Gorriz, E.: The relevance of deep chlorophyll maximum in the open Mediterranean Sea evaluated through 3D hydrodynamic-biogeochemical coupled simulations, Ecol. Model., 281, 26–37, 2014.

Mayot, N., D'Ortenzio, F., Ribera d'Alcalà, M., Lavigne, H., and Claustre, H.: Interannual variability of the Mediterranean trophic regimes from ocean color satellites, Biogeosciences, 13, 1901–1917, https://doi.org/10.5194/bg-13-1901-2016, 2016.

Mignot, Claustre, H., Uitz, J., Poteau, A., D'Ortenzio, F., and Xing, X.: Understanding the seasonal dynamics of phytoplankton biomass and the deep chlorophyll maximum in oligotrophic environments: A Bio-Argo float investigation, Glob. Biogeochem. Cycles, 28, 856–876, https://doi.org/10.1002/2013GB004781, 2014.

Mignot, A., Claustre, H., D'Ortenzio, F., Xing, X., Poteau, A., and Ras, J.: From the shape of the vertical profile of in vivo fluorescence to Chlorophyll-a concentration, Biogeosciences, 8, 2391–2406, https://doi.org/10.5194/bg-8-2391-2011, 2011.

Mignot, A., D'Ortenzio, F., Taillandier, V., Cossarini, G., and Salon, S.: Quantifying Observational Errors in Biogeochemical-Argo Oxygen, Nitrate, and Chlorophyll *a* Concentrations, Geophys. Res. Lett., 46, 4330–4337, https://doi.org/10.1029/2018GL080541, 2019.

Omand, M. M. and Mahadevan, A.: The shape of the oceanic nitracline, Biogeosciences, 12, 3273–3287, https://doi.org/10.5194/bg-12-3273-2015, 2015.

Osman, M. B., Das, S. B., Trusel, L. D., Evans, M. J., Fischer, H., Grieman, M. M., Kipfstuhl, S., McConnell, J. R., and Saltzman, E. S.: Industrial-era decline in subarctic Atlantic productivity, Nature, 569, 551–555, https://doi.org/10.1038/s41586-019-1181-8, 2019.

Park, J.-Y., Stock, C. A., Yang, X., Dunne, J. P., Rosati, A., John, J., and Zhang, S.: Modeling Global Ocean Biogeochemistry With Physical Data Assimilation: A Pragmatic Solution to the Equatorial Instability, J. Adv. Model. Earth Syst., 10, 891–906, https://doi.org/10.1002/2017MS001223, 2018.

Paulmier, A. and Ruiz-Pino, D.: Oxygen minimum zones (OMZs) in the modern ocean, Prog. Oceanogr., 80, 113–128, 2009.

Richardson, K. and Bendtsen, J.: Vertical distribution of phytoplankton and primary production in relation to nutricline depth in the open ocean, Mar. Ecol. Prog. Ser., 620, 33–46, https://doi.org/10.3354/meps12960, 2019.

Rousseeuw, P. J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math., 20, 53–65, https://doi.org/10.1016/0377-0427(87)90125-7, 1987.

Roxy, M. K., Modi, A., Murtugudde, R., Valsala, V., Panickal, S., Prasanna Kumar, S., Ravichandran, M., Vichi, M., and Lévy, M.: A reduction in marine primary productivity driven by rapid warming over the tropical Indian Ocean, Geophys. Res. Lett., 43, 826–833, https://doi.org/10.1002/2015GL066979, 2016.

Russell, J. L., Kamenkovich, I., Bitz, C., Ferrari, R., Gille, S. T., Goodman, P. J., Hallberg, R., Johnson, K., Khazmutdinova, K., and Marinov, I.: Metrics for the evaluation of the Southern Ocean in coupled climate models and earth system models, J. Geophys. Res. Oceans, 123, 3120–3143, 2018.

Salon, S., Cossarini, G., Bolzon, G., Feudale, L., Lazzari, P., Teruzzi, A., Solidoro, C., and Crise, A.: Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts, Ocean Sci., 15, 997–1022, https://doi.org/10.5194/os-15-997-2019, 2019.

Sauzède, R., Bittig, H. C., Claustre, H., Pasqueron de Fommervault, O., Gattuso, J.-P., Legendre, L., and Johnson, K. S.: Estimates of Water-Column Nutrient Concentrations and Carbonate System Parameters in the Global Ocean: A Novel Approach Based on Neural Networks, Front. Mar. Sci., 4, https://doi.org/10.3389/fmars.2017.00128, 2017.

Schartau, M., Wallhead, P., Hemmings, J., Löptien, U., Kriest, I., Krishna, S., Ward, B. A., Slawig, T., and Oschlies, A.: Reviews and syntheses: parameter identification in marine planktonic ecosystem modelling, Biogeosciences, 14, 1647–1701, https://doi.org/10.5194/bg-14-1647-2017, 2017.

Schmechtig, C., Poteau, A., Claustre, H., D'Ortenzio, F., and Boss, E.: Processing bio-Argo chlorophyll-A concentration at the DAC level, Ifremer, https://doi.org/10.13155/39468, 2015.

Schmechtig, C., Claustre, H., Poteau, A., and D'Ortenzio, F.: Bio-Argo quality control manual for the Chlorophyll-A concentration, Ifremer, https://doi.org/10.13155/35385, 2018.

Schmidt, H., Getzlaff, J., Löptien, U., and Oschlies, A.: Causes of uncertainties in the representation of the Arabian Sea oxygen minimum zone in CMIP5 models, Ocean Sci., 17, 1303–1320, https://doi.org/10.5194/os-17-1303-2021, 2021.

Schmidtko, S., Stramma, L., and Visbeck, M.: Decline in global oceanic oxygen content during the past five decades, Nature, 542, 335–339, https://doi.org/10.1038/nature21399, 2017.

Schneider, B., Bopp, L., Gehlen, M., Segschneider, J., Frölicher, T. L., Cadule, P., Friedlingstein, P., Doney, S. C., Behrenfeld, M. J., and Joos, F.: Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models, Biogeosciences, 5, 597–614, https://doi.org/10.5194/bg-5-597-2008, 2008.

Séférian, R., Bopp, L., Gehlen, M., Orr, J. C., Ethé, C., Cadule, P., Aumont, O., Salas y Mélia, D., Voldoire, A., and Madec, G.: Skill assessment of three earth system models with common marine biogeochemistry, Clim. Dyn., 40, 2549–2573, https://doi.org/10.1007/s00382-012-1362-8, 2013.

Steinacher, M., Joos, F., Frölicher, T. L., Bopp, L., Cadule, P., Cocco, V., Doney, S. C., Gehlen, M., Lindsay, K., Moore, J. K., Schneider, B., and Segschneider, J.: Projected 21st century decrease in marine productivity: a multi-model analysis, Biogeosciences, 7, 979–1005, https://doi.org/10.5194/bg-7-979-2010, 2010.

Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A. M., Rose, K. A., and Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems, J. Mar. Syst., 76, 4–15, https://doi.org/10.1016/j.jmarsys.2008.03.011, 2009.

Stramma, L., Johnson, G. C., Sprintall, J., and Mohrholz, V.: Expanding Oxygen-Minimum Zones in the Tropical Oceans, Science, 320, 655–658, https://doi.org/10.1126/science.1153847, 2008.

Tagliabue, A., Bopp, L., Dutay, J.-C., Bowie, A. R., Chever, F., Jean-Baptiste, P., Bucciarelli, E., Lannuzel, D., Remenyi, T., Sarthou, G., Aumont, O., Gehlen, M., and Jeandel, C.: Hydrothermal contribution to the oceanic dissolved iron inventory, Nat. Geosci., 3, 252–256, https://doi.org/10.1038/ngeo818, 2010.

Terzić, E., Lazzari, P., Organelli, E., Solidoro, C., Salon, S., D'Ortenzio, F., and Conan, P.: Merging bio-optical data from Biogeochemical-Argo floats and models in marine biogeochemistry, Biogeosciences, 16, 2527–2542, https://doi.org/10.5194/bg-16-2527-2019, 2019.

Thierry, V. and Bittig, H.: Argo quality control manual for dissolved oxygen concentration, 2018.

Thierry, V., Bittig, H., Gilbert, D., Kobayashi, T., Kanako, S., and Schmid, C.: Processing Argo oxygen data at the DAC level, Ifremer, https://doi.org/10.13155/39795, 2018.

Tuan Pham, D., Verron, J., and Christine Roubaud, M.: A singular evolutive extended Kalman filter for data assimilation in oceanography, J. Mar. Syst., 16, 323–340, https://doi.org/10.1016/S0924-7963(97)00109-7, 1998.

Tukey, J. W.: Exploratory Data Analysis, Addison-Wesley Publishing Company, 714 pp., 1977.

Ward, B. A., Friedrichs, M. A. M., Anderson, T. R., and Oschlies, A.: Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models, J. Mar. Syst., 81, 34–43, https://doi.org/10.1016/j.jmarsys.2009.12.005, 2010.

Westberry, T. K., Schultz, P., Behrenfeld, M. J., Dunne, J. P., Hiscock, M. R., Maritorena, S., Sarmiento, J. L., and Siegel, D. A.: Annual cycles of phytoplankton biomass in the subarctic Atlantic and Pacific Ocean, Glob. Biogeochem. Cycles, 30, 175–190, https://doi.org/10.1002/2015GB005276, 2016.

Williams, R. G. and Follows, M. J.: Ocean dynamics and the carbon cycle: Principles and mechanisms, Cambridge University Press, 2011.

Wong, Keeley, Robert, Carval, Thierry, and Argo Data Management Team,: Argo Quality Control Manual for CTD and Trajectory Data, https://doi.org/10.13155/33951, 2015.