Combining Neural Networks and Data Assimilation to enhance the spatial impact of Argo floats in the Copernicus Mediterranean biogeochemical model

Carolina Amadio[1], Anna Teruzzi[1], Gloria Pietropolli[1,2], Luca Manzoni[2], Gianluca Coidessa[1], and Gianpiero Cossarini[1]

[1]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale – OGS, Trieste, 34100, Italy

[2]Dipartimento di Matematica e Geoscienze, Università degli Studi di Trieste H2bis Building, Via Alfonso Valerio 12/1, 34127 Trieste, Italy;

Correspondence: C. Amadio (camadio@ogs.it)

Abstract. ~~Biogeochemical-Argo (BGC-Argo~~Biogeochemical Argo (BGC Argo) float profiles provide substantial information for key vertical biogeochemical dynamics and are successfully integrated in biogeochemical models via data assimilation approaches. Although results on the BGC-Argo assimilation are encouraging, data scarcity remains a limitation for their effective use in operational oceanography.

To address availability gaps in the BGC-Argo profiles, an Observing System Experiment (OSE), that combines Neural Network (NN) and Data Assimilation (DA), has been performed here. NN was used to reconstruct nitrate profiles starting from oxygen profiles and associated Argo variables (pressure, temperature, salinity), while a variational data assimilation scheme (3DVarBio) has been upgraded to integrate ~~BGC-Argo~~ BGC Argo and reconstructed observations in the Copernicus Mediterranean operational forecast system (MedBFM). To ensure high quality of oxygen data, a post-deployment quality control method has been developed with the aim of detecting and eventually correcting potential sensors drift.

The Mediterranean OSE features three different setups: a control run without assimilation; a multivariate run with assimilation of BGC-Argo chlorophyll, nitrate, and oxygen; and a multivariate run that also assimilates reconstructed observations.

The general improvement of skill performance metrics demonstrated the feasibility in integrating new variables (oxygen and reconstructed nitrate). Major benefits have been observed in reproducing specific BGC process-based dynamics such as the nitracline dynamics, primary production and oxygen vertical dynamics.

The assimilation of BGC-Argo nitrate corrects a generally positive bias of the model in most of the Mediterranean areas, and the addition of reconstructed profiles makes the corrections even stronger. The impact of enlarged nitrate assimilation propagates to ecosystem processes (e.g., primary production) at basin wide scale, demonstrating the importance of the assimilation of BGC-profiles in ~~complementing satellite ocean colour assimilation~~forecasting the biogeochemical ocean state.

1 Introduction

The ~~Array for Real-time Geostrophic Oceanography (Argo)~~ Argo programme appears to be one of the better examples of countries and human resource capacities in working together to provide global data coverage (Miloslavich et al., 2019) that supports the investigation of the present (analysis), future (forecast) and past years (reanalysis) ocean state conditions. In the last 10 years, the increase of ~~in-situ~~ in situ observations from autonomous platforms (Johnson et al. 2013 and Johnson and Claustre 2016) has opened up new perspectives for biogeochemical oceanographers. Indeed, ~~BGC-Argo (biogeochemical Argo ,~~ BGC Argo (Argo 2023) has yielded new insights in describing the interior of the global ocean (Le Traon, 2013) and key processes such as the deep chlorophyll maximum (Mignot et al. 2014, Barbieux et al. 2019, D'ortenzio et al. 2020, Ricour et al. 2021, and Barbieux et al. 2022), ~~oxygen and (Capet et al., 2016),~~ nutrients vertical fluxes (Taillandier et al. 2020 and Wang et al. 2021b)~~and~~ , carbon exports (Dall'Olmo and Mork 2014 and Wang and Fennel 2023) and oxygen dynamic (Capet et al., 2016).

~~Among the BGC sensors,~~ With approximately 270,000 profiles worldwide (as of July 2023), oxygen (O2) is currently the most ~~common measured variable, with approximately 270,000 profiles worldwide until now, which~~ commonly measured variable. The count of O2 profiles is double that of suspended particles and chlorophyll, and more than four times ~~those~~ that of nitrate, downwelling irradiance, and pH (https://biogeochemical-argo.org). ~~Currently, the~~ Since 2019, the availability of nitrate and chlorophyll profiles has progressively decreased due to the high cost of the sensor (Dall'Olmo personal communication). The number of oxygen profiles instead decreased initially (2019-2022), but since 2022 is stable or slightly increasing. In the future, Argo Italy envisages mounting oxygen sensors on all Argo floats in the Mediterranean Sea (Discussion in the workshop on "Copernicus Marine requirements for the in situ Observing System", 14-15 September 2023).

The BGC-Argo data are distributed by the Global Data Assembly Centres (GDACs, e.g., Coriolis, NOAA) ~~distributes oxygen data~~ in Real Time (RT) Adjusted (AM) and Delayed Mode (DM). The quality of AM data is controlled within 24 hours using internationally agreed and automatic quality-control ~~procedures performed at the surface, along the entire vertical profiles and along the trajectory (Thierry and Bittig, 2021).~~ (QC) procedures, while DM data are generally distributed a few months later (nearly six months) in more rigorous form (Li et al., 2020). The QC tests, conducted across all the data mode levels, aim to assign a quality flag to every observation. Data labeled as 1, 2, 5, and 8 are categorized as good, probably good, changed, and interpolated value, respectively. The flag 9 indicates missing data, while flags 3 and 4 denote data as probably bad or bad.

~~Major efforts have been devoted to improve~~ In the case of oxygen, the QC mainly perform at the surface, along the entire vertical profiles and along the trajectory (Thierry and Bittig, 2021), excluding specific tests at depth. The implementation of O2 QC tests is mainly devoted to improving the long-term reliability and accuracy of autonomous ~~O2 measurements (Sauzède et al., 2017) .~~ measurements (Sauzède et al., 2017) in particular with respect to the sensor drift (the optode drift).

When the sensor drift exists, it is higher in the storage, out of the water, than during the deployment. As described in Takeshita et al. (2013) and in Maurer et al. (2021), raw oxygen data from floats ~~can have~~ may exhibit errors of up to 20% in terms of oxygen saturation (at the surface) due to sensor drift ~~during storage out of the water. By improving the accuracy up to 5-10%, 1st-order correction methods can correct storagedrift~~ occurring during the storage. This drift is generally corrected by multiplying the oxygen concentrations ~~with~~ for a gain factor term that is derived from a reference dataset (Johnson et al., 2015). ~~Additionally, oxygen measurements are calibrated using values of saturation at the surface (in water for the older ones and in air for floats with new sensors, Bushinsky et al. 2016). Despite correction and calibration progress~~Despite this correction can improve the accuracy up to 5-10%, Maurer et al. (2021) and Bushinsky et al. (2016) found a drift in about 25% (~~mean drift -0.07~~with a mean of -0.07% per year~~) and 70 % (mean -0.12 %~~ , a standard deviation of 0.65%, and a total range of 1.1 to 1.2% per year) and 70% of analyzed floats, respectively. ~~Positive and negative values of drift were~~ Given the logistical challenges in recovering deployed floats, an in situ (or during deployment) drift >1% per year can be likely observed (Bushinsky et al., 2016) . The drift can be both positive or negative as found in Johnson and Claustre (2016) and Bittig et al. (2018b). ~~Therefore, the~~

The development and dissemination of a post-deployment ~~quality control (QC) method has been encouraged~~ oxygen QC aims to avoid spurious results (Wang et al., 2020) and to distinguish between ocean signals or trends (e.g., deoxygenation) ~~and potential optode drift.~~ from potential drifts. This allows to obtain more robust datasets suitable for specific numerical modelling applications.

~~The combination of in-situ observations and numerical models represents a promising approach to exploit the BGC-Argo potentialities. Indeed, BGC-Argo data are used for several BGC modelling tasks such as: (i) model tuning (Wang et al., 2020) , (ii)~~ Aiming at optimally combining observations and model information to obtain a closer description of reality, the data assimilation (DA) underpins decades of progress in ocean prediction (Geer, 2021). On one hand, progresses began with an increase in the number of available observations over the past decade encompassing both the number of measured variables and the total observations used for model tuning (Wang et al. 2020, Yumruktepe et al. 2023 and Wang and Fennel 2023) and validation (Terzić et al. 2019, Salon et al. 2019 and Wang et al. 2021a)~~, and (iii) assessment of the BGC ocean state and variability through data assimilation (DA) (Cossarini et al. 2019and Teruzzi et al. 2021~~ . On the other hand DA scheme were progressively updated to enable multivariate and multiplatform assimilation (Cossarini et al. 2019, Teruzzi et al. 2021 and d'Ortenzio et al. 2021)~~. DA underpins decades of progress in ocean prediction (Geer, 2021) by increasing the type and number of observational datasets and~~ , retrieve associated uncertainty into ~~a prediction model~~prediction models, and solving problems connected to uneven distribution and/or scarcity of the observations (Buizza et al., 2022).

~~Given their capacity~~ In recent years, neural network (NN) algorithms have been increasingly used to solve and analyze specific tasks related to observation processing and DA. The main strength of NN algorithms lies in their ability to approximate continuous functions (Hornik et al., 1989) ~~, NN algorithms match specific DA tasks such as~~ in remarkably low computational times. For these reasons, DA techniques have been recently augmented with NN-based tools, e.g., for: bias correction (Kumar et al. 2015 and Zhou et al. 2021), ~~cross calibration (Lary et al., 2018),~~ reformulation of observation operators (Storto et al., 2021), cross calibration and new product creation or dataset reconstruction (Lary et al., 2018). As an example, ocean ~~color~~ colour (OC) datasets were employed to test Multi-Layer Perceptrons (MLP, namely the most common NN) by retrieving past and long-term BGC time-series of phytoplankton and chlorophyll (Martinez et al. 2020a, Martinez et al. 2020b, Roussillon et al. 2023). Moreover, in Sauzède et al. (2016), MLP serves to infer chlorophyll vertical BGC distribution from OC. High performance in predicting biogeochemical states (e.g., oxygen) from physical profiling floats measurements were achieved in Stanev et al. (2022) for the Black Sea.

In Sauzède et al. (2017), ~~an~~ a MLP-NN is used to approximate nutrient concentration and carbonate system from physical ~~and~~ Argo and BGC-Argo oxygen profiles, and the updated version of Bittig et al. (2018c) allows refining the previous work with the so-called Canyon-b NN method. A configuration to adapt global Canyon-b NN in the Mediterranean Sea region is developed by Fourrier et al. (2020). A further update of the application of the MLP method in the Mediterranean Sea is provided in Pietropolli et al. (2023), by achieving a lower error in the nutrients predictions through a larger training dataset, a hyperparameter refinement and a two-step quality control of the input data. Given its potential in predicting nutrient profiles, the MLP-NN model outputs are valuable datasets that can be used to fill the gap in the availability of in situ observations in data assimilation.

In the context of operational oceanography, the biogeochemical modelling component of the Copernicus Marine Service for the Mediterranean Sea (MedBFM) provides analysis, short term forecast (Salon et al., 2019) and long term reanalysis (Cossarini et al., 2021), including the assimilation of satellites ~~Ocean Colour (OC)~~ OC and BGC-Argo observations (Salon et al., 2019).

~~The~~ In the MedBFM, the 3DVarBio variational assimilation scheme, ~~3DVarBio, of MedBFM~~ has evolved over time by including a greater number of observation types and variables. Starting from the first release that included OC data assimilation in the open ocean (Teruzzi et al., 2014), the assimilation has progressively ~~included~~ developed to handle coastal OC observations (Teruzzi et al., 2018), chlorophyll and nitrate profiles from ~~BGC-Argo~~ BGC Argo (Cossarini et al. 2019 and Teruzzi et al. 2021 respectively). ~~Given~~ Considering the growing availability of O2 from ~~BGC-Argo, in this paper we propose~~ BGC Argo, this paper presents an additional upgrade of the MedBFM to include BGC-Argo oxygen assimilation, with a novel post-deployment quality control, and the integration of NN reconstructed profiles in the assimilation scheme.

The constant evolution of the observation networks and assimilation capacities requires an updated understanding of the impact of observation on the numerical model result (Gasparin et al., 2019). This can be achieved by using the numerical assimilative models in Observing System Experiments (OSEs) where the impact of existing observations on the model performance is assessed (Le Traon et al., 2019).

In this paper, ~~our~~ the OSE experiment, ~~that~~ which combines data assimilation and neural network in a ~~sequential modular approach(hereafter NN-MLP-MED), aims at quantifying how~~ modular approach, aims to quantify how the Argo and BGC-Argo ~~networks~~ network can be exploited. The sequential use of the NN and DA schemes provides flexibility in using one module independently of the other, depending on the needs of the overall system (Buizza et al., 2022). The DA module used in this work is the 3DVarBio data assimilation scheme described in Teruzzi et al. 2021 and updated to assimilate oxygen BGC-Argo profiles. The NN module is the NN-MLP described in Pietropolli et al., 2023 for the Mediterranean Sea (hereafter NN-MLP-MED).

Spatial and temporal impacts of the OSE have been evaluated using classic and new skill performance metrics in three two-year (2017-2018) numerical experiments performed using the MedBFM coupled with the 3DVarBio: a control run (HIND) without assimilation; a multivariate run (DAfl) with assimilation of BGC-Argo chlorophyll, nitrate, and oxygen; and a multivariate run that also assimilates ~~in-situ~~ in situ observations and reconstructed ones (DAnn).

~~Given~~ Because of its particular characteristics and the ~~high~~ density of BGC-Argo profiles, the Mediterranean Sea represents an ideal site for OSE experiments to evaluate the potentiality of the BGC-profiles assimilation.

The Mediterranean Sea is an anti-estuarine semi-enclosed sea characterized by specific physical and biogeochemical dynamics (Pinardi et al., 2015), with a complex horizontal circulation consisting of mesoscale and sub-basins scale gyre structures, transitional cyclonic and anticyclonic gyres and eddies, influenced by bathymetric features interconnected by currents and jets (Oddo et al., 2009). Despite its relatively limited extent in the mid-latitude temperate zone, the Mediterranean Sea has a considerable BGC variability that can be roughly approximated in an oligotrophic West-East gradient with low nutrient availability at the surface, insufficient to sustain significant phytoplankton biomass (Siokou-Frangou et al. 2010 and Marañón et al. 2021) and a deeper nitracline in the east (>120m) with respect to the west (<100m). Additionally, chlorophyll has a particular seasonal cycle with pronounced winter/early spring surface blooms only in the western part and a few locations in the eastern part. During summer, a deep chlorophyll maximum follows the stratified and oligotrophic conditions at increasing depth moving eastward (>100m at East and <100m in the West) (Teruzzi et al., 2021). Dissolved oxygen has a subsurface maximum at about 50m, with higher values in the west (partly due to the dependence of oxygen solubility on temperature). Noticeable differences are observed in the intermediate layers where the oxygen minimum ranges between 300 (west) and 1000 m (east) (Di Biagio et al., 2022).

The paper is organized as follows. After a brief presentation of the OSE approach, each component and the experimental setup are described in detail (Section 2). In the following section (Section 3), we describe the results of the novel NN-MLP-MED and the assimilation simulations by using different skill metrics to assess model capability in reproducing the main biogeochemical seasonal dynamics. A discussion of some key issues involved in the NN and DA is provided in Section 4, then the paper closes with some final remarks (Section 5).

2 Methods

A novel combined Neural Network (NN-MLP-MED) and Data Assimilation (3DVarBio) approach is included in the Mediterranean MedBFM model system to integrate BGC-Argo and reconstructed profiles into biogeochemical simulations of the Mediterranean Sea.

Our OSE experiment is based on a sequential modular approach (Buizza et al., 2022) consisting of a post-deployment quality control method of O2, hereafter QC O2 procedure, a trained multi-layer perceptron NN (Pietropolli et al., 2023) and a data assimilation scheme (the 3DVarBio variational scheme of MedBFM, Figure 1).

The input of the first two modules, QC O2 and NN-MLP-MED are the BGC-Argo and Argo datasets, while the final 3DVarBio module also takes the enhanced datasets as input: quality checked O2 (qcO2) and reconstructed nitrate (recNO3, Figure 1).

In the following sections, the novel modules of the MedBFM system (i.e., the QC O2 procedure and the NN-MLP-MED scheme) and the dataset (BGC-Argo and reconstructed datasets) are described together with the revised ~~3D-VarBio~~3DVarBio.

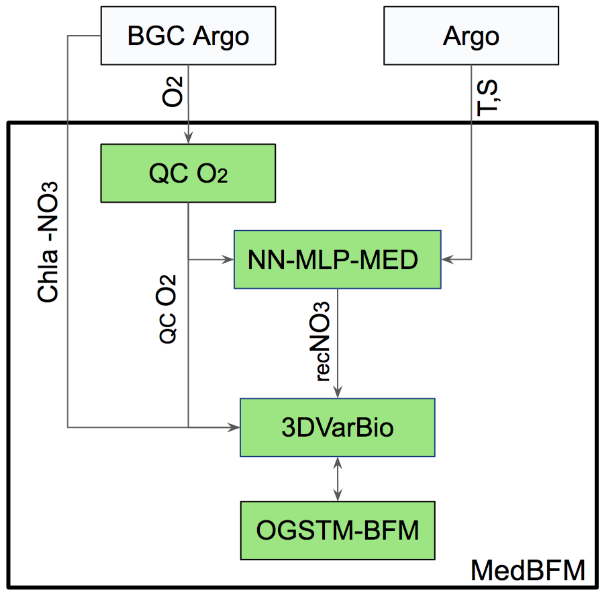

Figure 1. Flowchart of the NN-MLP-MED and DA approach. In green boxes: the modules. In plain boxes: the datasets. Arrows refers to Argo (temperature and salinity) and BGC Argo profiles of chlorophyll (Chla), oxygen (qcO_2) nitrate (NO_3) and reconstructed nitrate ($recNO_3$)

2.1 The regional model for the Mediterranean Sea, MedBFM

The MedBFM consists of the OGS transport model (OGSTM) based on the OPA 8.1 system (Foujols et al., 2000) and updated according to the Lazzari et al. (2012) and Lazzari et al. (2016) versions, the BFM, Biogeochemical Flux Model described in Vichi et al. (2007a) and Vichi et al. (2007b), and the 3DVarBio variational assimilation scheme as in Teruzzi et al. (2014) and Teruzzi et al. (2018).

OGSTM solves advection, diffusion, sinking terms and the free surface and variable volume-layer effects on the transport of tracers (Salon et al., 2019), and it is forced by the output (current, T, S and sea surface height) of the NEMO3.2 model. OGSTM and NEMO3.2 share the same bathymetry and z* grid configuration, open boundary and river conditions (Coppini et al., 2023).

The Biogeochemical Flux Model, BFM, is a biomass and functional group based marine ecosystem model. BFM solves governing equations for nine living-organic state variables (diatoms, autotrophic nanoflagellates, picophytoplankton, dinoflagellates, carnivorous and omnivorous mesozooplankton, bacteria, heterotrophic nanoflagellates, and microzooplankton, macronutrients (nitrate, phosphate, silicate and ammonium) and labile, semi-labile, and refractory organic matter and oxygen. In addition, the BFM includes a carbonate system model (Cossarini et al. 2015a and Canu et al. 2015).

2.2 3DVarBio data assimilation scheme

Based on 3DVarBio (Teruzzi et al. 2014, Teruzzi et al. 2018, Cossarini et al. 2019 and Teruzzi et al. 2021), the assimilation ~~scheme~~ module adopted in the present work integrates oxygen, chlorophyll and nitrate to update all the assimilated variables as well as all the phytoplankton biomasses and phosphate.

The 3DVarBio is a variational data assimilation scheme (Teruzzi et al., 2014) based on the minimization of a cost function *(J)* which relies on the misfit between the model background (x_b) and the observations (y) weighted with the respective error covariance matrices (*B* and *R*) as follows:

$$J(x_a) = (x_a - x_b)^T B^{-1}(x_a - x_b) + (y - H(x_b))^T R^{-1}(y - H(x_b)) \tag{1}$$

Here, the observation operator (*H*) maps the values of model background state in the observation space. Following Dobricic et al. (2006), the background error covariance matrix, *B*, is factorised as B=VV^T with V=$V_V V_H V_B$. The V operators describe different aspects of the error covariances: the vertical error covariance (V_V), the horizontal error covariance (V_H), and the state variable error covariance (V_B). V_V is defined by a set of reconstructed profiles evaluated by means of an Empirical Orthogonal Function (EOF) decomposition applied to a validated multi-annual 1998-2015 run (Teruzzi et al., 2018). EOFs are computed for 12 months and 30 coastal and open sea sub-regions in order to account for the variability of biogeochemical anomaly fields. V_H is built using a Gaussian filter whose correlation radius modulates the smoothing intensity. ~~A~~ As in Cossarini et al. (2019), in this work the correlation radius is non-uniform~~and~~ , direction-dependent~~correlation radius has been implemented as in Cossarini et al. (2019)~~, and ranges between 12 and 20 km (16 km on average). V_B operator consists of prescribed monthly and sub-region varying covariances among the biogeochemical variables (e.g., nitrate to phosphate). Specifically, for the assimilation of chlorophyll, the V_B operator includes a balance scheme that maintains the ratio among the phytoplankton groups and preserves the physiological status of the phytoplankton cells (i.e., ~~preserve optimal values for the internal~~ preserves the internal ratios between the chlorophyll, carbon and nutrients ~~quota).~~ as described in Teruzzi et al. 2014).

The operators V_V and V_B of the 3DVarBio have been updated for the assimilation of oxygen. V_V involved the calculation of specific EOF profiles for oxygen including a localization function to avoid ~~spurious assimilation~~ unrealistic corrections due to possible spurious error covariances in the deepest part of the water column.

V_B included only a new direct relation for oxygen (i.e., oxygen assimilation update only the oxygen itself), given that it has been shown that it barely affects other variables (Skakala et al., 2021). In the BFM model equations, few formulations depend on oxygen concentration (e.g., nitrification). Indeed, when the euphotic zone of the open ocean is well oxygenated, oxygen dynamics has a limited impact on the biogeochemical cycles.

Assimilated observations are composed by the QC BGC-Argo listed in Table 1. Oxygen and nitrate profiles in the 0-600 m layer are used in the assimilation, while chlorophyll is assimilated in the 0-200 m layer.

The observation error covariance matrix R is diagonal with a monthly varying error in chlorophyll (Cossarini et al., 2019). In both nitrate and reconstructed nitrate profiles, the observation error ~~is~~ remains constant in time and increases along the vertical~~with a constant values of~~ . Within the 0-450 m layer the error is set at $0.25\ \mathrm{mmol\ m^{-3}}$ ~~in the 0-450 meters layer~~

~~(Mignot et al., 2019) and linearly increasing of~~ as in Mignot et al. and then linearly increases up to 0.35 $\mathrm{mmol\ m^{-3}}$ between ~~450-600m (nitrate~~ 450 and 600 m (the maximum assimilation depth). This adjustment aims to avoid inconsistencies between the deeper (below 600 m) and the lower part of the assimilated layer. Although the accuracy of the reconstruction of profiles is 0.87 $\mathrm{mmol\ m^{-3}}$ (Pietropolli et al., 2023), we decided to not use different values of error for the two nitrate subsets in order to show the highest potential impact of the OSE.

Observation error for oxygen is set to 5 $\mathrm{mmol\ m^{-3}}$ in the upper 200 meters of depth and gradually goes to 20 $\mathrm{mmol\ m^{-3}}$ in correspondence of the maximum assimilation depth. These values correspond to the uncertainty associated with the oxygen dataset described in Feudale et al. (2022)

2.3 The ~~neural network architecture~~ Architecture of the Neural Network module and the ~~reconstructed nitrate dataset~~Reconstructed Nitrate Dataset

The NN-MLP-MED (Pietropolli et al., 2023) is the evolution of previous MLP architectures developed to predict low-sampled variables (e.g., nutrients) starting from high-sampled ones (e.g., temperature) (Sauzède et al. 2017 , Bittig et al. 2018c and Fourrier et al. 2020).

NN-MLP-MED is a deterministic Feed-Forward Neural Network based on a MLP structure. It consists of the merging of 10 different MLP architectures, each one with the same input and output features, composed by two hidden layers with varying numbers of neurons per layer. The final prediction resulting from the NN-MLP-MED is the mean of all the predictions of these components. The data flow of the MLP-based approach follows the forward direction from the input to the output layers through the neurons which compose the layers. In our OSE experiment, the trained NN-MLP-MED reconstructs nitrate profiles from temperature and salinity (Argo), oxygen (BGC Argo) and float date, latitude and longitude.

The NN-MLP-MED presents some novel elements with respect to the mentioned methods (and in particular with respect to Canyon-Med in Fourrier et al. 2020), which lead to improved results. Firstly, the input dataset includes a larger sample size and broader coverage of the Mediterranean Sea region~~. Secondly, the quality of the~~ , i.e., the quality controlled EMODnet2018_int data collection that integrates the in situ aggregated EMODnet data collections (Buga et al., 2018) and the observations listed in Lazzari et al. (2016) and Cossarini et al. (2015b).

Secondly, the input dataset benefits from a two-step quality check process, removing noisy and unreliable samples. The neural network architecture was also modified to enhance prediction performance by ~~incorporating nonlinear functions, adjusting neuron count,~~ accurately selecting a performing nonlinear function, adjusting and optimizing the ~~training algorithm. The error of reconstructed nitrate, obtained by using the EMODnet as validation dataset, was 0.5 (Pietropolli et al., 2023)~~

~~In our OSE experiment, the trained~~ amount of neurons for each layer of the MLP model, and choosing a different optimization strategy to train the algorithm. NN-MLP-MED ~~is used to reconstruct nitrate profiles from temperature and salinity (Argo), oxygen (BGC-Argo) and float date, latitude and longitude. The reconstruction~~ also includes a vertical smoothing step (running mean of 5-10 m window) and ~~an adjustment to the~~ a climatological adjustment at depth (600 m~~climatology derived from EMODnet~~) derived from the EMODnet dataset (Salon et al., 2019).

~~The input nitrate dataset for assimilation is made up of 938~~ Uncertainty of reconstructed nitrate associated to the EMODnet validation dataset is 0.5 $\mathrm{mmol\ m^{-3}}$, while it reaches 0.87 $\mathrm{mmol\ m^{-3}}$ when it predicts the BGC-Argo ~~profiles and~~ dataset (Pietropolli et al., 2023).

After incorporating the reconstructed profiles (recNO3), the nitrate dataset used for the assimilation expands to 2146 ~~reconstructed nitrate~~ profiles from the initial 938 nitrate (NO3) profiles (Table 1). ~~The reconstructed nitrate profiles are located 61% in the western and 39% in the eastern Mediterranean Sea , thus providing a larger and more homogeneous spatial coverage as shown in Figures 2~~Generated by the NN-MLP-MED module, the reconstructed dataset offers a broad spatial coverage across the 16 regions of the Mediterranean Sea (Figure 2) as well as a balanced distribution of nitrate data throughout both winter and summer seasons (Figure 3).

2.4 BGC-Argo data and the post-deployment ~~oxygen quality control~~QC O2 module

BGC-Argo profiles ~~available between the~~ from 2017-2018 ~~period~~ were downloaded from the Coriolis GDAC (~~last visit on~~ Argo 2022, last visited in July 2022). ~~Adjusted and delayed mode data were selected~~ We collected both AM and DM data for oxygen and chlorophyll. For nitrate we selected DM data, while AM data were incorporated after undergoing correction via Canyon-b NN method or using the World Ocean Atlas (WOA18) collection (Garcia et al., 2019) as explained in Johnson et al. (2021). For the three variables we use data flagged as good, probably good, changed and interpolated values (flags 1,~~while exclusively DM data were considered for nitrate.~~ 2, 5 and 8).

Table 1 reports the total number of BGC-Argo profiles, characterized by a significant number of oxygen and chlorophyll data against the relative paucity of nitrate. Figure 2 shows the spatial distribution of ~~the~~ BGC profiles of chlorophyll and nitrate ~~, while~~ across the Mediterranean Sea. The oxygen coverage can be approximated by merging nitrate and reconstructed nitrate profiles locations. ~~Despite the lack of data in specific~~

To provide more clarity in analyzing the data availability, the Mediterranean Sea has been divided into 16 sub-basins~~(Alboran~~ :

- in the Western Mediterranean Sea: Alboran Sea (alb), South Western ~~Ionian and Northern Adriatic Seas), all~~ Mediterranean west (swm1), South Western Mediterranean east (swm2), North Western Mediterranean (nwm), Northern Tyrrhenian (tyr1) and Southern Tyrrhenian (tyr2).

- in the Eastern Mediterranean Sea: Northern Adriatic (adr1), Southern Adriatic (adr2), Western Ionian (ion1), Eastern Ionian (ion2), Northern Ionian (ion3), Western Levantine (lev1), Northern Levantine (lev2), Southern Levantine (lev3) Eastern Levantine (lev4) and Aegean Sea (aeg).

All the three BGC variables have a fairly homogeneous spatial coverage between the ~~western and eastern Mediterranean Sea.~~ Western and Eastern Mediterranean Sea, except for few sub-basins not covered (alb, ion1 and adr1 see Figure 2) , and a generally 5-day temporal sampling frequency. Higher sampling frequencies (< 5 days) are registered for the 20% of profiles.

Since oxygen sensors may drift and lose accuracy over time, the accurate determination of dissolved oxygen is typically more challenging and requires some form of correction (Johnson et al., 2015). Expressed in % per year, the loss of accuracy is

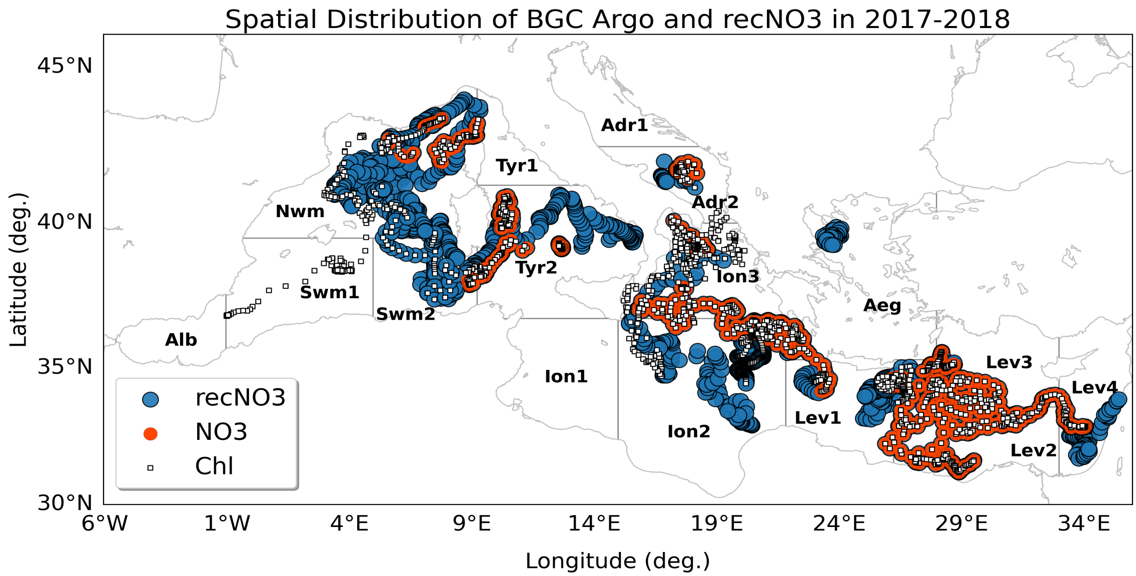

Figure 2. BGC-profiles of ~~chlorophyll-a~~ chlorophyll (~~red~~Chl, in white), ~~Nitrate in-situ~~ nitrate in situ (~~orange~~NO3, in red) and reconstructed ~~Nitrate~~ nitrate (~~grey~~recNO3, in blue) assimilated in Mediterranean Sea (2017-2018). Subdivision of the Mediterranean domain in sub-basins used for the validation. According to data availability and to ensure consistency and robustness of the metrics, different subsets of the sub-basins or some combinations among them can be used for the different metrics: lev=lev1+lev2+lev3+lev4; ion=ion1+ion2+ion3; tyr=tyr1+tyr2; adr=adr1+adr2; swm=swm1+swm2.

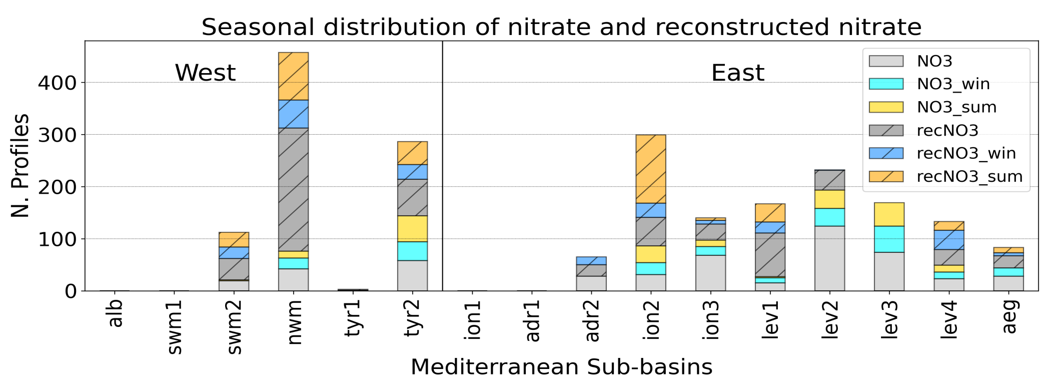

Figure 3. Nitrate and reconstructed nitrate profiles seasonal availability. Light gray (autumn and spring), cyan (winter), and yellow (summer) bars represent the availability of nitrate in situ data (used in the run called DAfl). Gray (autumn and spring), light blue (winter), and orange (summer) striped bars indicate the availability of reconstructed nitrate (used in the run called DAnn).

observed over the time, particularly 12 months after the deployment (https://www.euro-argo.eu). Deep ocean drift is considered as a proxy for oxygen sensor drift because of the lack of seasonal and annual signals for oxygen at depth (Takeshita et al., 2013).

Here, the optode ~~sensor in-situ~~ drift is evaluated through non parametric ~~tests~~ methods (RANSAC and ~~Theil Sen~~Theil-Sen) at two different depths (600 and 800 ~~meters)~~ m) to avoid possible fake drift detection because of changes in the water masses. Tests are applied when the life of a float is longer than ~~1 year.~~ one year. Conversely, if the available float time series is less than one year, the profiles are not corrected because the float lifetime is considered too short to account for in situ sensor drift.

~~The RANSAC and Theil Sen methods split~~ Used for linear and non-linear regression problems, the RANSAC and Theil-Sen methods automatically partition the oxygen dataset into ~~a set of~~ inliers and outliers~~and drift is estimated only using inliers, avoiding possible biases due to the outliers~~ . In order to avoid possible biases (Dang et al. 2008 and Fischler and Bolles 1981), these methods calculate the drift based on the data subset identified as inliers.

~~The~~ In our approach, the presence of a drift is established when all four drift estimates (RANSAC at 600 and 800 m, Theil-Sen at 600 and 800 m) agree in sign and their average value ~~is greater than~~ (D_avg) exceeds $1\ \mathrm{mmol\ m^{-3}\ y}$. This threshold ~~was~~ is chosen on the basis of results in Bittig et al. (2018b).

~~When detected, the suspicious~~ Subsequently, the identified drift is removed from the oxygen profiles. This is achieved by setting the ~~computed drift values (i.e., the average of the four estimates)~~ D_avg at 600 meters and linearly interpolating toward the surface, where drift is set equal to zero. ~~Indeed, it can be assumed that O2 values at surface are already fixed~~ As highlighted by Thierry and Bittig (2021), there is a lack of specific tests at depth, although several tests are performed near the surface by the GDACs~~(Thierry and Bittig, 2021).~~ . The presence of near-surface tests motivates our decision to mitigate the correction's impact at the surface.

2.5 Design of numerical experiments

Three numerical experiments ~~were~~ are performed to analyze the impact of different assimilation setups. ~~Simulated~~ The simulated period is 1.1.2017-31.12.2018, and the MedBFM module setup mostly corresponds to the standard adopted in the Mediterranean Analysis and Forecast biogeochemical system of the ~~Marine Copernicus Service. Set up includes~~Copernicus Marine Service. This set up includes: open boundary conditions in the Atlantic~~,~~ ; climatological input of nutrients, carbon and alkalinity for 39 rivers and the Dardanelles Straits; ~~and~~ initial conditions from EMODnet dataset (~~Details~~ details are provided in Salon et al. 2019); and a 3-years spin up using the 2017 forcings in perpetual mode.

Our experimental setup differs from the standard one for the physical forcing, which are from Mediterranean Copernicus reanalysis (Escudier et al., 2021), and the initial conditions of oxygen which are retrieved from BGC-Argo float climatology computed after QC O2 procedure (described in section 2.4).

The three simulations, which share the same setup except for the assimilated datasets, are: (1) control run without assimilation (HIND); (2) assimilation of BGC-Argo chlorophyll, nitrate and oxygen ~~and~~ (DAfl) and (3) assimilation of additional reconstructed nitrate profiles used to enhance the DAfl assimilative set up (DAnn).

Before integrating data in the 3D-VarBio, the same pre-assimilation ~~assessments~~ assessment described in Teruzzi et al. (2021) ~~were applied for chlorophyll~~ is applied to the chlorophyll profiles. Nitrate profiles are rejected if concentration at the surface is higher than $3\ \mathrm{mmol\ m^{-3}}$. ~~Finally, oxygen~~ At surface, the oxygen profile exclusion is evaluated ~~on the basis of the difference with the oxygen saturation values, using a~~ by calculating the difference between the uppermost oxygen measurement

Test Case	Chl	O2	NO3	Updated variables
HIND	–	–	–	–
DAfl	1773	1924	938	phyto biomass, NO3 , O2 and PO4
DAnn	1773	1924	2146	phyto biomass, NO3 , O2 and PO4

Table 1. Summary of the numerical experiments and assimilated BGC-profiles

and the oxygen saturation (derived from temperature and salinity data from the Argo dataset as in Garcia et al. 2019). Profiles are excluded when this difference reaches the threshold of 10 $\mathrm{mmol\ m^{-3}}$~~and comparing oxygen data at 600m with respect to a reference dataset using a~~ . At 600 meters, the difference between oxygen and a climatological reference oxygen at depth is calculated. Profiles are excluded when the difference reaches the threshold of 2 times the standard deviation of the same reference dataset. ~~For the~~ As reference dataset, we ~~used~~ chose the EMODnet2018_int data collection that integrates the ~~in-situ~~ in situ aggregated EMODnet data (Buga et al., 2018)) and the datasets listed in Lazzari et al. (2016) and Cossarini et al. (2015b). The EMODnet2018_int dataset is available for 16 sub-basins ~~(see Figure 2)~~ in the Mediterranean Sea (Figure 2).

During the data assimilation, profiles ~~can be excluded when model-observation misfit is higher than given thresholds~~are excluded when innovation exceeds specific threshold rules. For chlorophyll, the threshold is set at $2\mathrm{mg\ m^{-3}}$~~and it must be found in at least 5 vertical levels in the 0-50m layer~~. For nitrate, the ~~misfit thresholds are set to~~ thresholds are 2 and 3 $\mathrm{mmol\ m^{-3}}$ ~~in~~ for the 0-50 m and ~~250-600m~~ 250-600 m layers, respectively ~~. Exceeding misfit has to be found in at least 5 vertical levels. Oxygen profiles are discarded by defining misfit thresholds of~~ (as in Teruzzi et al. 2021). Oxygen thresholds are 30 and 50 $\mathrm{mmol\ m^{-3}}$ for the 0-150 m and 150-600 m layers respectively (thresholds are roughly 3 times the standard deviation of the climatology computed on EMODnet data for the different sub-basins). Exceeding values have to be found in at least 5 vertical levels ~~in the 0-150m and 150-600m layersrespectively.~~

within the specified layers. These exclusions aim to prevent corrections that could trigger unstable dynamics after the assimilation (Teruzzi et al. 2021, Storto et al. 2011, Sakov and Sandery 2017 and Waller et al. 2018). The excluded profiles range from 0.1% for chlorophyll to less than 1% for nitrate.

3 Results

3.1 The ~~oxygen~~ post-deployment ~~quality check method~~QC O2 module

The ~~post deployment oxygen QC method allowed to automatically correcting in-situ~~ product of our QC O2 module is a QCed dataset available at https://zenodo.org/records/10391759.

The QC O2 module allowed for the automatic correction of in situ sensor drifts.

~~Of~~ Out of the 40 floats available ~~in the~~ between 2017-2018~~period, the drift analysis was applied to .~~, we conducted drift analysis on 16 floats, while 24 floats ~~could not be analyzed~~ remained unanalyzed due to the limited length of the timeseries.

~~Over the~~ Among these 16 floats, we identified a significant drift ~~was found~~ in 13~~of them~~: 4 ~~were affected by~~ with a positive drift and 9 ~~by~~ with a negative one. The remaining 3 floats had ~~lower drift values than~~ a drift values below the prescribed threshold (Section 2.4).

~~The~~ At a depth of 600 meters, the absolute average correction ~~of~~ for the 13 floats is about 4.3 $\mathrm{mmol\ m^{-3}}$ y~~performed at 600 meters of depth. This quantity is in line .~~. This value aligns with the ranges expressed in terms of sensor drift percentage in Bittig et al. (2018a) (1-1.5%).

Figure 4 shows ~~an example of~~ the evolution of oxygen profiles ~~of~~ for a quasi-stationary float (6902687) ~~detected for a~~ following the application of drift correction. ~~In line with the conclusion reported in several works such as~~ Consistent with findings in various studies (e.g., Bittig et al. (2018a) and Maurer et al. (2021)~~, the presence of a drift , like the one detected by our protocol, may reveal the~~), the detection of drift by our QC O2 suggests a possible tendency of the ~~oxygen sensor~~ optode to slowly degrade over time. After 2 years, the bias due to the drift ~~was~~ reaches approximately 5 $\mathrm{mmol\ m^{-3}}$ (1st December 2017 profiles in Figure 4~~.)~~).

~~After removing of drift , the deep oxygen concentrations results to be~~ The removal of drift brings the oxygen concentration at 600 m closer to the EMODnet climatological data ~~, allowing to include a higher number of profiles (example~~ (as exemplified in Figure 4, green star). This leads us to infer that our drift correction enables the inclusion of more profiles in the assimilated ~~O2 datasets(Figure 1)~~oxygen datasets.

3.2 Validation using Satellite and BGC-Argo datasets

~~Skills performance~~ Skill performances of the simulations listed in Table 1 are evaluated by comparing model results with satellite ~~OC~~ Copernicus OC product (i.e., OCEANCOLOUR_MED_BGC_L3_MY_009_143 from marine.copernicus.eu, last visited in July 2023) of chlorophyll and BGC-Argo profiles ~~. While for the satellite comparison the model daily averages are considered, the model~~ (Argo, 2022). The satellite comparison used daily model output. The model first guess (i.e., the model state at 1pm before the assimilation) is ~~used for~~ instead used for the metrics based on BGC-Argo ~~.~~ profiles. While the use of the first guess is a common practice in DA applications (Hollingsworth et al., 1986), it is worth to remind that this comparison should be considered as a semi-independent validation, given that two consecutive profiles of the same BGC-Argo float can share a certain degree of correlation.

The Root Mean Square Error (RMSE) metric is evaluated ~~in~~ during winter (from February to April, FMA) and summer (from June to August, JJA) 2017 and 2018 to investigate the model's ~~capacity to reproduce the~~ capability to reproduce specific bloom and stratification conditions ~~in~~ within 16 ~~Mediterranean Sea~~ sub-basins of the Mediterranean Sea (described in Section 2.4 and in Figure 2) or in an aggregated combination of them~~(Figure 2) .~~. This latter includes six macro-basins: the South Western Mediterranean Sea (Swm) consisting of swm1 and swm2; the North Western Mediterranean (Nwm/NWM) represented solely by the nwm; the Tyrrhenian Sea (Tyr), consisting of tyr1 and tyr2; the Ionian Sea (Ion), consisting of ion1, ion2, and ion3; the Adriatic Sea (Adr), consisting of adr1 and adr2; and the Levantine Sea (Lev), consisting of lev1, lev2, lev3 and lev4.

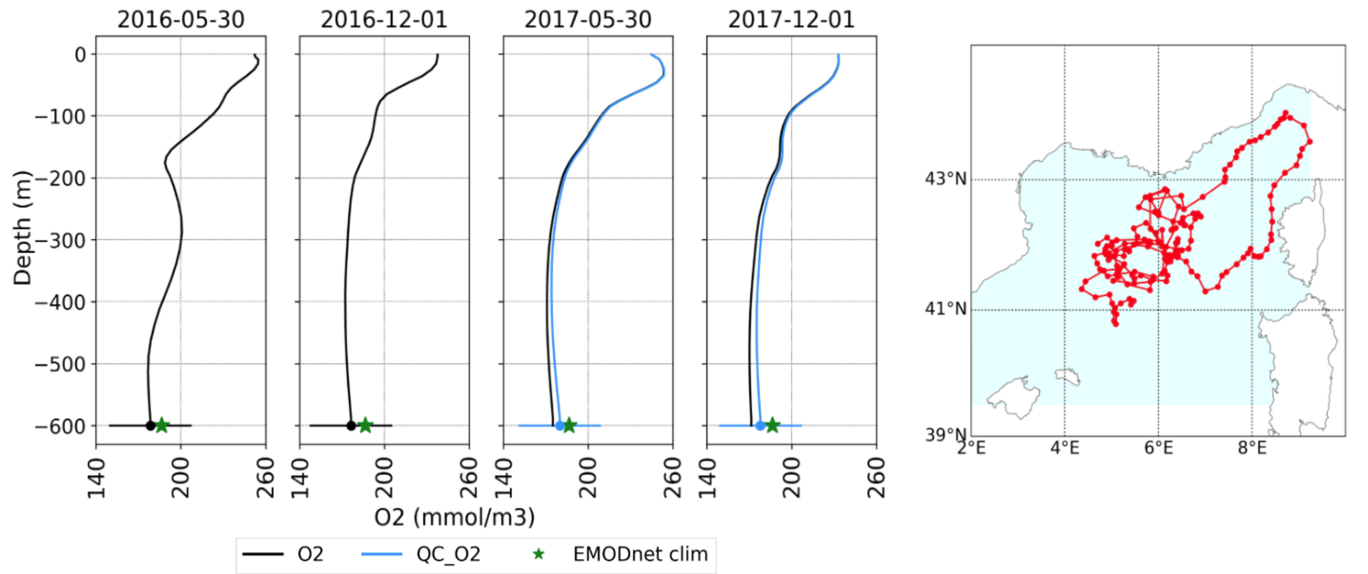

Figure 4. ~~Original~~ Depiction of original (black) and corrected (blue) oxygen profiles ~~of~~ for float 6902687 ~~on~~ across four selected dates. ~~EMODnet climatology for nwm subbasin is reported (~~The green star ~~)~~refers to the EMODnet O2 climatological value in the nwm sub-basin and the horizontal line to the EMODnet O2 standard deviation at 600 m

~~Satellite~~ The satellite OC L3 products ~~from~~ downloaded from the Copernicus Marine Service catalogue ~~were~~ are interpolated from 1 km to the model resolution and a composite weekly average was computed to ensure gap-free maps, as in Teruzzi et al. (2014).

The Winter RMSE ~~with respect to~~ concerning the OC chlorophyll in HIND spans between ca. 0.09 to 0.21 $\mathrm{mg\ m^{-3}}$ with a maximum in the ~~Alboran Sea~~ alb region (Figure 5). The ~~addition~~ inclusion of multivariate DA (~~DAfl) has a positive impact with a reduction of surface errors of~~ in DAfl) positively impact the model performances, reducing surface errors by 6.5% mainly observed in the eastern sub-basins. A further reduction of RMSE (up to 10%) with respect to HIND is then obtained with DAnn ~~showing~~ highlighting that enlarging the nitrate float network leads to improvements in reproducing surface phytoplankton dynamics. ~~All~~ Except for alb and swm1, where no nitrate data (in situ and reconstructed) were available, all the Mediterranean sub-basins ~~show an RMSE reduction with the exception of alb, swm and nwm .~~ exhibit a reduction in RMSE during winter. In the nwm, the RMSE in the DAfl assimilative setup is higher than in the HIND run. However, in DAnn (light-blue striped bar of nwm in Figure 3) the enlarged nitrate dataset positively affects the chlorophyll dynamics at surface.

A generalized slight worsening in the assimilated runs can generally be observed during the summer stratification period and especially the Eastern sub-basins. The RMSE with respect to OC chlorophyll, which increases in all sub-basins, is fairly similar in the two assimilation runs: about 6% and 7.5% in DAfl and DAnn, respectively. ~~It should be noted that the~~ Despite the introduction of a significant number of reconstructed nitrates in some sub-basins (e.g., orange stripped lines of nwm and ion2 in

Figure 3), the inclusion of recNO3 profiles does not positively impact summer chlorophyll shallow statistics. The RMSE values in summer are an order of magnitude lower than in winter, reflecting the seasonal chlorophyll variability in the Mediterranean Sea (i.e., the very low values of chlorophyll at the surface).

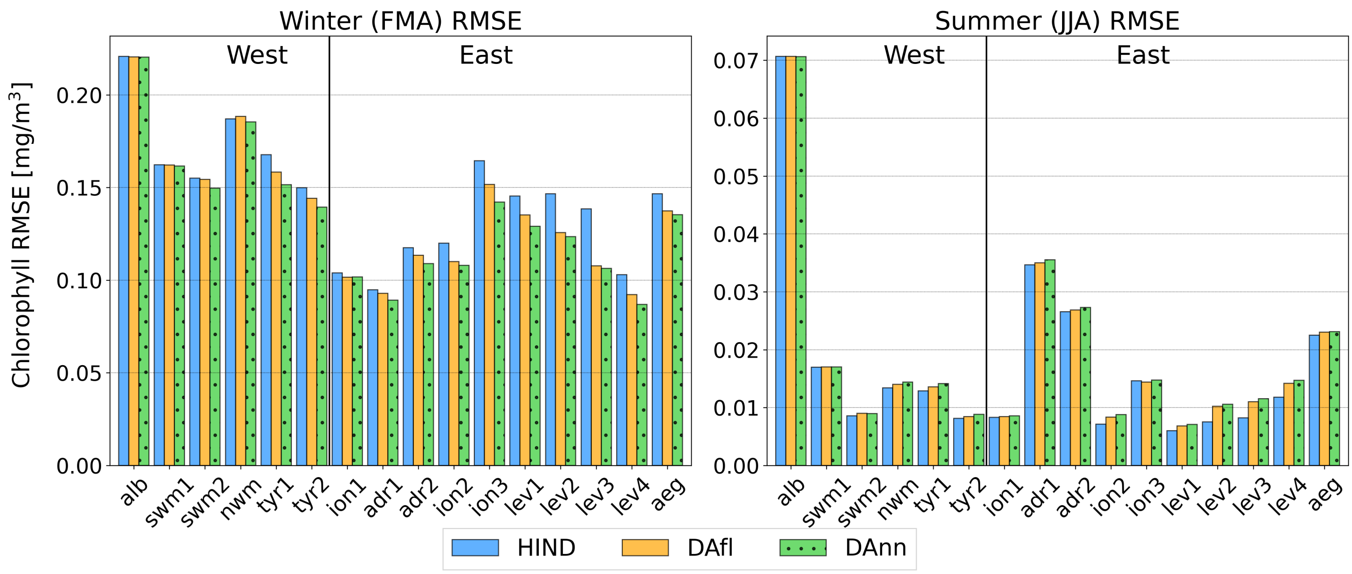

Figure 5. Seasonal chlorophyll RMSE: Winter bloom and Summer stratification seasons in the Mediterranean Sea sub-basins for the HIND run (light blue), the DAfl run (in orange) and the DAnn run (~~green~~dotted-green). The black vertical line represents the subdivision of the Mediterranean Sea in West and East sectors.

The RMSE metrics based on BGC-Argo are computed for the six selected aggregated macro-basins ~~(Alboran, South West Mediterranean, North West Mediterranean, Tyrrhenian, Ionian and Levantine Seas) and~~ and in selected layers (~~0-10m, 10-30m, 30-60m, 60-100m, 100-150m, 150-300m and 300-600m~~0-10 m, 10-30 m, 30-60 m, 60-100 m, 100-150 m, 150-300 m and 300-600 m) and are shown in Figure 6 for nitrate (top panel), chlorophyll (middle panel) and oxygen (bottom panel). ~~It is worth recalling that only in-situ profiles are used in the validation (i.e., reconstructed NN profiles are used only for assimilation~~ The statistics computed over the aggregate basin provide a more robust results (e.g., they are computed over a larger number of profiles) even if possible spatial patterns of the errors can be damped. Thus, this choice might limit the analysis on whether/how different nitrate assimilation setups affect chlorophyll and oxygen dynamics (see Section 3.3).

As expected, the assimilation of ~~in-situ~~ in situ BGC-Argo considerably improves the quality of modelled nitrate with respect to the HIND run~~(Figure 6). Winter RMSE reduction goes from~~. During winter, the average RMSE reduction is 40% ~~(DAfl) to approximately~~ in DAfl, and increases to 46% ~~, when also reconstructed profiles are assimilated, while the reduction of summer RMSE increases from~~ in DAnn, while in summer the average reduction reaches 59% in DAfl ~~to~~ and 63% in DAnn ~~. Maximum~~ (first row in Figure 6). The most significant RMSE reduction of ~~RMSE of DAnn with respect~~ DAnn compared to DAfl is observed in ~~nwm and tyr (winter) and in ion (summer)~~ Nwm and Tyr (0-450 m) during winter, and in Ion (0-100 m) in summer. This impact can be directly ascribed to the ~~increased number of reconstructed nitrate in these sub-basins and seasons where~~ profiles availability (Figure 3) and additional profiles generate more persistent corrections.

~~The advantages of assimilating chlorophyll profiles have been already shown in Teruzzi et al. (2021). Here, improvements related to~~ Since the DAfl and DAnn simulation share the same chlorophyll assimilative setup, the RMSE improvements in terms of chlorophyll assimilation can be ~~observed in nwm, ion and lev in winter and at depth in tyr, ion and lev in~~ evaluated comparing the HIND with the DAfl or DAnn simulations (Figure 6 middle panel). We observe slight enhancements in simulating chlorophyll in Nwm (0-100 m) and Lev (0-200 m) during winter and in Tyr, Ion and Lev (50-200 m) during summer (Figure 6 middle panel). Even if phytoplankton dynamics depend on nutrients dynamics, the positive impact of DAnn on nitrate RMSE does not transfer to the the vertical chlorophyll statistics in the DAnn. This is ~~mainly because the DAfl and DAnn simulations assimilate the same chlorophyll dataset, and the~~ because the direct chlorophyll assimilation is more effective than the dynamical model adjustment after nitrate and reconstructed nitrate assimilation in the areas close to the observed chlorophyll profiles.

Assimilating oxygen profiles ~~enable~~ enables reducing the model-BGC floats RMSE by about 30% during winter and summer. ~~The~~ In winter, the correction involves the whole water columns ~~with a maximum correction between 150-600m during winter in the west and along the entire profiles in the east (ion and lev) in both winter and summer. The addiction~~ in the East (Lev and Ion, third row in Figure 6) and deeper layers (150-600 m) in the West (Swm, Nwm) and Adr. In summer, the impact is mainly observed in Tyr, Ion and Lev. As discussed in Section 2.2, the integration of reconstructed profiles in the ~~DAfl run~~ DAnn simulation does not significantly ~~affect the quality of oxygen .~~ impact the oxygen dynamics. Finally, it is important to note that whenever a reconstructed nitrate is assimilated, oxygen is also assimilated.

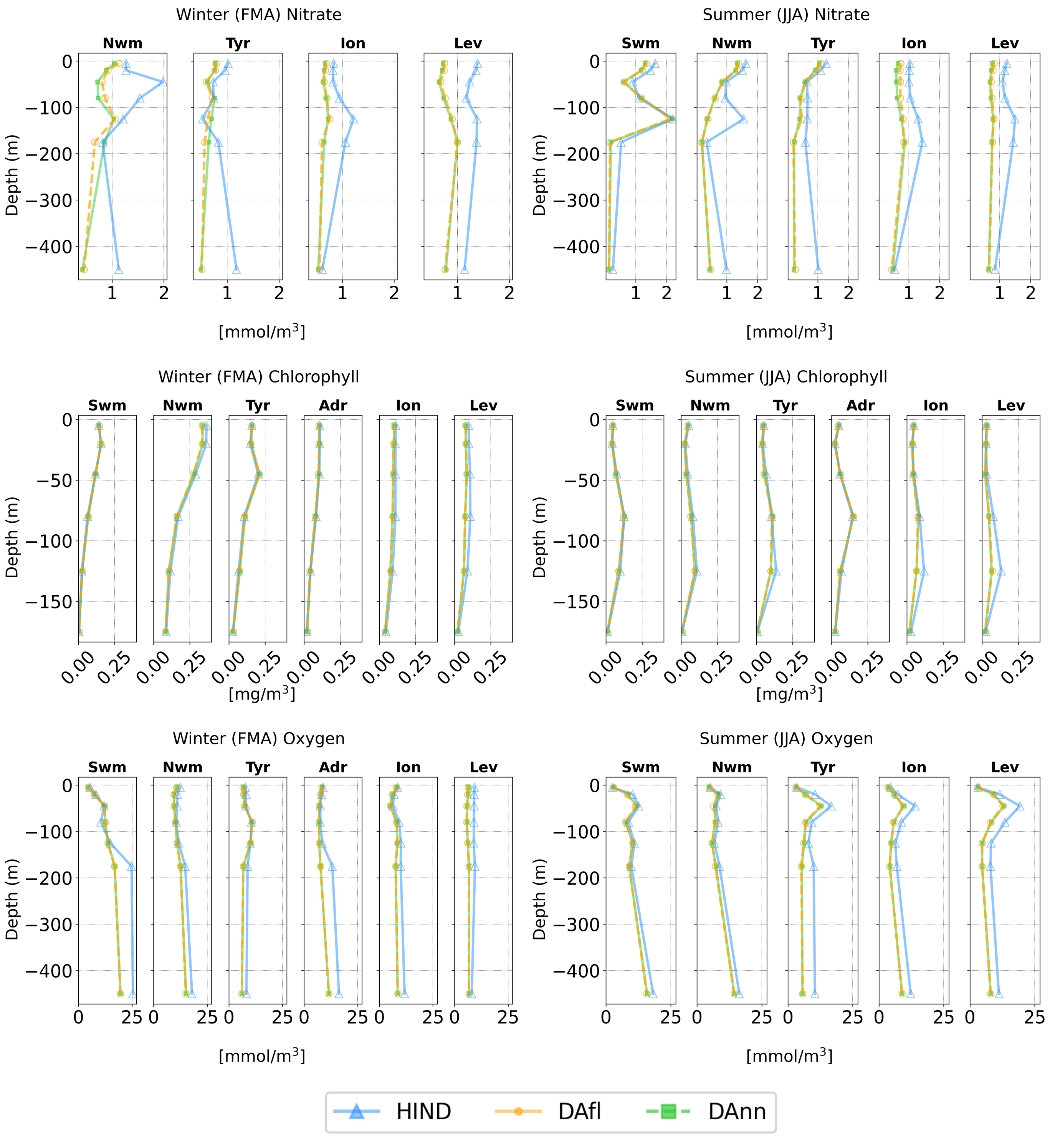

Figure 6. Seasonal Nitrate, Chlorophyll and Oxygen profile of RMSE (top, middle, bottom): Bloom (left) and Stratification (right) seasons in the Mediterranean Sea aggregated sub-basins for the HIND run (pale blue), DAfl run (orange) and DAnn (green)

3.3 Integration of ~~NN~~ NN-MLP-MED and DA modules: the impact

3.3.1 Impacts on biogeochemical vertical dynamics

~~As expected, the profile assimilation plays a major role~~ To assess the impact of profile assimilation in changing the vertical gradients of biogeochemical variables~~. Figures 9,~~ , Figures 7, 8, 9 and 10 show the ~~impact of the assimilation in two~~ Hovmöller diagrams of the spatial averages of nitrate, phosphate, chlorophyll and oxygen for two selected sub-basins ~~where the number of reconstructed profiles is high (~~(first and second columns for Nwm and ion2 with boundaries indicated in the map of Figure 2) and ~~in the~~ for the entire Mediterranean Sea (third column). ~~The two sub-basins represent two different~~ This representation offers additional details on the vertical impact of the reconstructed nitrate profile assimilation with respect to the validation of Figure 6 that considers only model points corresponding to the location of BGC-Argo profiles. Nwm and ion2 represent distinct trophic conditions in the Mediterranean Sea and are also characterized by high number of assimilated reconstructed nitrate profiles (Figure 3). The North Western Mediterranean ~~(nwm)~~ has higher level of nutrient ~~concentration~~ concentrations and more intense surface ~~bloom~~ blooms in winter (Siokou-Frangou et al. 2010 and Di Biagio et al. 2022). ~~In summer, nwm has~~ During summer, Nwm exhibits a shallow nitracline, higher chlorophyll concentration at the deep chlorophyll maximum (DCM) ~~, shallow nitracline,~~ and shallow subsurface oxygen maximum (SOM) (first column in Figures ~~9,~~ 7, 8, 9 and 10). ~~On the contrary, more oligotrophic conditions and~~ Conversely, the eastern sub-basin is characterized by a deeper nitracline and DCM ~~are found in the eastern subbasin~~ and more oligotrophic conditions (ion2, second column of Figures ~~9,~~ 7, 8, 9 and 10). ~~The assimilation of nitrate~~

Considering nitrate, the multivariate assimilation (DAfl) corrects a general positive bias of the model in all the Mediterranean areas (blue pattern in Figure 7). The addition of reconstructed profiles makes the corrections stronger. ~~At the Mediterranean scale~~On average, the nitrate concentration below the nitracline (the depth at which nitrate concentration is 2 $\mathrm{mmol\,m^{-3}}$) decreases by 8% and 11% in DAfl and DAnn runs, respectively. ~~Nitracline depth changes (i.e. deeper values) by DAfl assimilation by a few (nwm) and tens of meters (ion2). The deepening of the nitracline becomes more intense with the inclusion of reconstructed profiles (DAnn). The differences~~ Both the assimilation runs also exhibit changes of the nitracline depth with more intense deepening in the DAnn simulation. Differences between the assimilation and ~~reference runs accumulates over timeand eventually reach a stationary phase in~~ the reference run accumulate over time. The rate of this accumulation is highest during the first year and decreases during the second year~~in the~~. These differences remain almost constant in sub-basins with a high number of BGC-Argo and reconstructed profiles ~~, such as nwm and ion2~~(e.g., Nwm in Figure 7). On the other hand, considering the ion2 and the whole Mediterranean Sea, which comprises some under-sampled areas (e.g., ~~southern Ionian and southern western basin~~ion1 and ion3), the effect of DA corrections is still propagating after the two years (third column of Figure 7).

Very similar patterns are also observed in the Hovmöller diagrams of phosphate (Figure 8), which is an updated variable of the multivariate variational assimilation scheme through ~~nitrate~~nitrate-phosphate covariance. In fact, the general negative corrections on phosphate fields are linked to the high positive values of the covariance matrix between nitrate and phosphate (Teruzzi et al., 2021).

~~As a consequence of both the direct assimilation of chlorophyll profiles and the dynamical model adjustment after nitrate assimilation~~Considering chlorophyll (Figure 9), the main ~~effects of DAfl are to slightly reduce the intensity of chlorophyll concentration in the DCM~~ difference between DAfl and HIND is a slight reduction of the DCM chlorophyll concentration (e.g., variation smaller than 5% with respect to HIND simulation) and ~~in adjusting~~ a correction of the timing of the surface winter blooms (second row in Figure 9). Even if the chlorophyll validation (Figure 6) ~~has not shown~~ does not show significant differences between DAfl and DAnn, the basin wide averages of DAnn display more intense corrections with respect to DAfl in terms of DCM depth and chlorophyll intensity and overall chlorophyll concentration (~~figure~~ Figure 9). Over the 0-200 m layer of the ~~entire~~ whole Mediterranean Sea, the chlorophyll decreases with ~~respects~~ respect to HIND are 4% and 5% for DAfl and DAnn, respectively.

~~BGC-Argo oxygen profiles~~ Corrections on oxygen dynamics after the multivariate assimilation (DAfl, second row in Figure 10) ~~provides~~ are either positive or negative ~~corrections~~ depending on the ~~observation-model bias which varies in time and space (e.g., mostly negative in nwm and mostly positive in~~ area and the period of the year. In particular, corrections are mostly positive in ion2~~). On a~~ , while the Nwm sub-basin shows negative corrections in the subsurface layer and positive ones in the upper layer of the second year. On the Mediterranean basin-wide scale, the average correction is ~~of~~ 0.2% for the ~~0-200m~~ 0-200 m layer. The addition of the nitrate reconstructed profiles does not alter the correction pattern with an average correction of 0.3%. ~~The only noticeable difference~~ However, the largest differences between the two assimilation runs can be spotted in areas with a high density of reconstructed profiles during summer (e.g., ~~nwm~~Nwm, first column in Figure 10). As observed in the nitrate and chlorophyll ~~figures~~Hovmöller diagrams, the assimilation of reconstructed profiles causes a decrease of the summer productivity in the DCM layer. Consequently, less oxygen is produced generating the negative changes in the DCM layer in the bottom left panel of Figure 10. Because of the smaller amount of subsequent sinking organic matter, less oxygen is consumed in remineralization processes in layers below the DCM in late summer and autumn, and positive oxygen changes are generated, particularly during 2018.

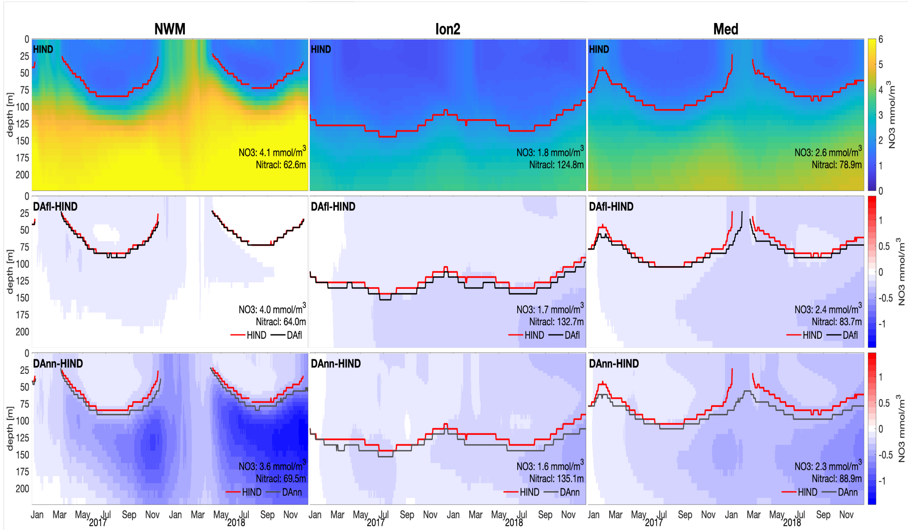

Figure 7. Hovmöller diagram of ~~chlorophyll~~ nitrate of ~~hindcast~~ HIND simulation (first row) and differences between assimilation runs and ~~hindcast~~ HIND (second and third rows) for 2 sub-basins (nwm and ion2) and the Mediterranean Sea (med). Evolution of the depth of nitracline (the depth at which nitrate concentration is 2 mmol m^{-3}) for the three runs: red (~~hind~~HIND) and black (DAfl and DAnn) lines. The averages of 0-200m concentration and of nitracline for the simulated period are reported.

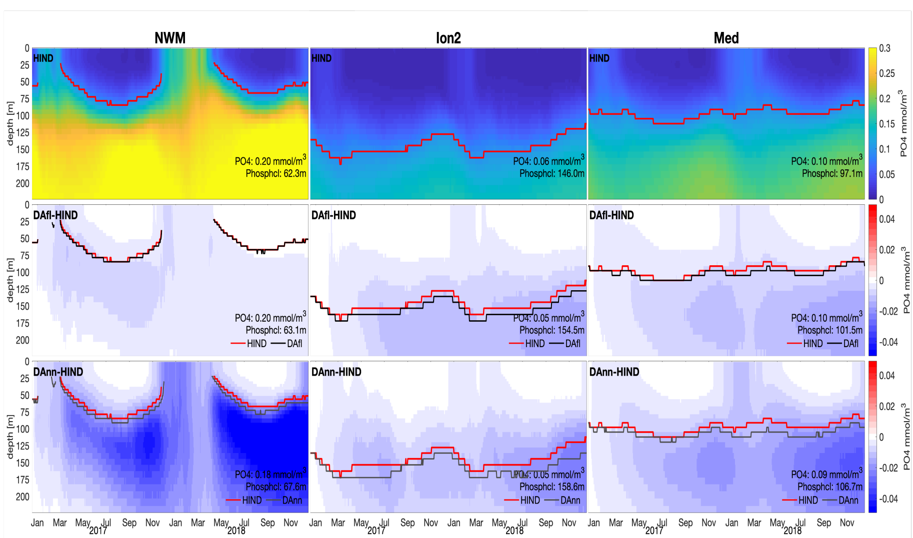

Figure 8. Hovmöller diagram of ~~nitrate~~ phosphate of ~~hindcast~~ HIND simulation (first row) and differences between assimilation runs and ~~hindcast~~ HIND (second and third rows) for 2 sub-basins (nwm and ion2) and the Mediterranean Sea (med). Evolution of the depth of ~~nitracline~~ phosphocline (the depth at which phosphate concentration is 0.1 mmol m^{-3}) for the three runs: red (~~hind~~HIND) and black (DAfl and DAnn) lines. The averages of 0-200m concentration and of ~~nitracline~~ phosphocline for the simulated period are reported.

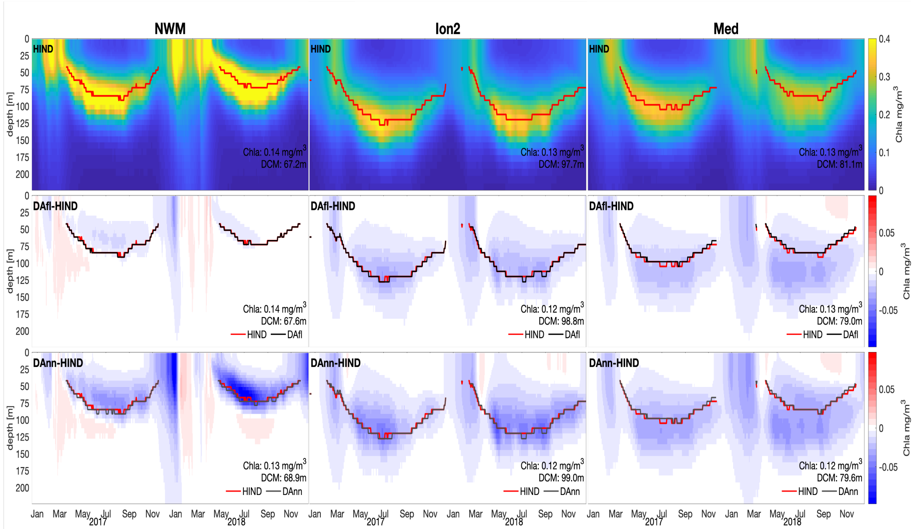

Figure 9. Hovmöller diagram of ~~phosphate~~ chlorophyll of ~~hindcast~~ HIND simulation (first row) and differences between assimilation runs and ~~hindcast~~ HIND (second and third rows) for 2 sub-basins (nwm and ion2) and the Mediterranean Sea (med). Evolution of the depth of ~~phosphocline~~ Deep Chlorophyll Maximum (DCM) for the three runs: red (~~hind~~HIND) and black (DAfl and DAnn) lines. The averages of 0-200m concentration and of ~~phosphocline~~ nitracline for the simulated period are reported.

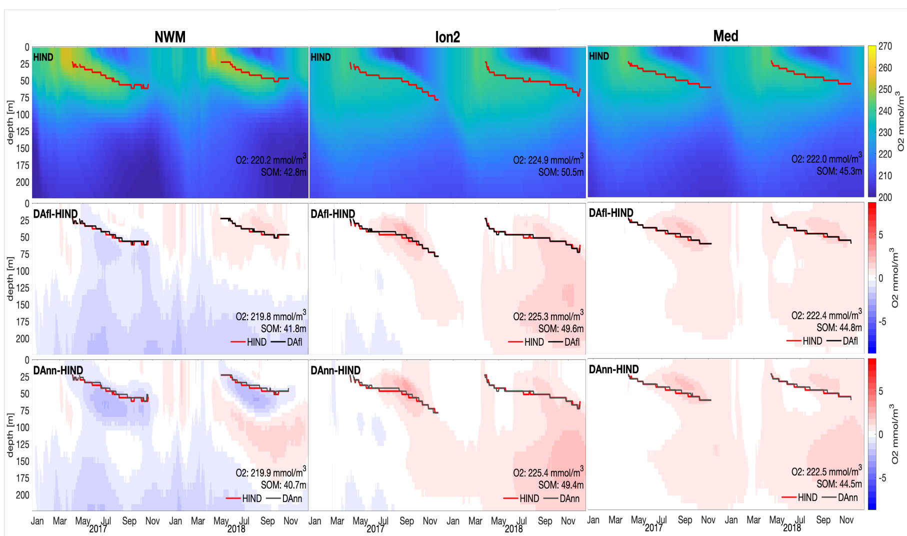

Figure 10. Hovmöller diagram of oxygen of ~~hindcast~~ HIND simulation (first row) and differences between assimilation runs and ~~hindcast~~ HIND (second and third rows) for 2 sub-basins (nwm and ion2) and the Mediterranean Sea (med). Evolution of the depth of subsurface oxygen maximum (SOM) for the three runs: red (~~hind~~HIND) and black (DAfl and DAnn) lines). The averages of 0-200m concentration and of SOM for the simulated period are reported.

3.3.2 Impact on ecosystem indicator (net primary production)

Net primary production (NPP) integrates phytoplankton growth and respiration processes which are at the basis of the marine trophic food web. ~~Assimilation~~ The assimilation of chlorophyll and nitrate together with the updates of phosphate, directly and indirectly affect primary production, ~~since they impact~~ as they influence both phytoplankton biomass and nutrient availability. Thus, the comparison of primary production among the three simulations reveals how the assimilation impacts on a key indicator that integrates several marine ecosystem processes. Seasonal maps of net primary production integrated over the 0-200 m layer in the HIND, DAfl and DAnn simulations (Figure 11) confirm that the assimilation's impact varies spatially and temporally.

In the DAfl simulation, the ~~largest~~ most significant differences in primary production ~~with respect~~ compared to the HIND simulation are located in the ~~eastern Mediterranean~~ Eastern Mediterranean Sea with a decrease of NPP of nearly 10% in the Levantine ~~sub-basins~~ macro-basin and in the Ionian Sea close to the Greek coast ~~. Reductions are slightly larger in~~ (first and second row of Figure 11). This reduction is particularly pronounced during winter. In the ~~western~~ Western Mediterranean the impacts on primary production are negligible in both seasons with ~~the only exception of a~~ a slight reduction (5%~~reduction~~) in winter in the Tyrrhenian Sea. ~~Areas with changes on NPP corresponding to the areas with assimilation of float profiles that include nitrate. In the DAnn simulation , the~~

The DAnn simulation shows more pronounced impacts on primary production ~~are more intense than in DAfl and~~ compared to the DAfl simulation (second and third rows of Figure 11). The main differences between the DAnn and DAfl are highlighted by the black contour line in Figure 11 (differences larger than 15 $\mathrm{mgC\ m^{-2}\ d^{-1}}$). Specifically, during winter, a decrease in NPP is mainly observed in the ~~impacted areas are larger~~Nwm, Ion, and Tyr, while in summer the reductions in NPP is observable in the Nwm and Ion. ~~In particular, primary production is decreased also in areas such as the western Mediterranean in summer (8%) and in the northwestern Mediterranean and in the Tyrrhenian Sea in winter (5%).~~

~~Moreover, a further reduction of NPP occurs in the Ionian Sea in both seasons~~As shown in Figure 3, lev1 and lev4 exhibit a considerable abundance of reconstructed nitrate profiles in both winter and summer seasons. This enhanced profile availability results in spatially localized increases of impacts. The absence of recNO3 data and the presence of NO3 data in lev2 and lev3 seem to damp the eventual propagation of impacts across lev2 and lev3.

In general, the impact on primary production is greater where nitrate observations or nitrate reconstructed observations are assimilated (Figure 3), suggesting a dynamical bottom-up control of primary production. In fact, the weaker fertilization of the surface layer in DAnn, which occurs for both macronutrients after assimilation (Figure 7 and 8), ~~appears to be the main cause of reduced NPP, outweighing the effects due to changes in phytoplankton biomass after chlorophyll assimilation~~causes a reduction of the net primary production.

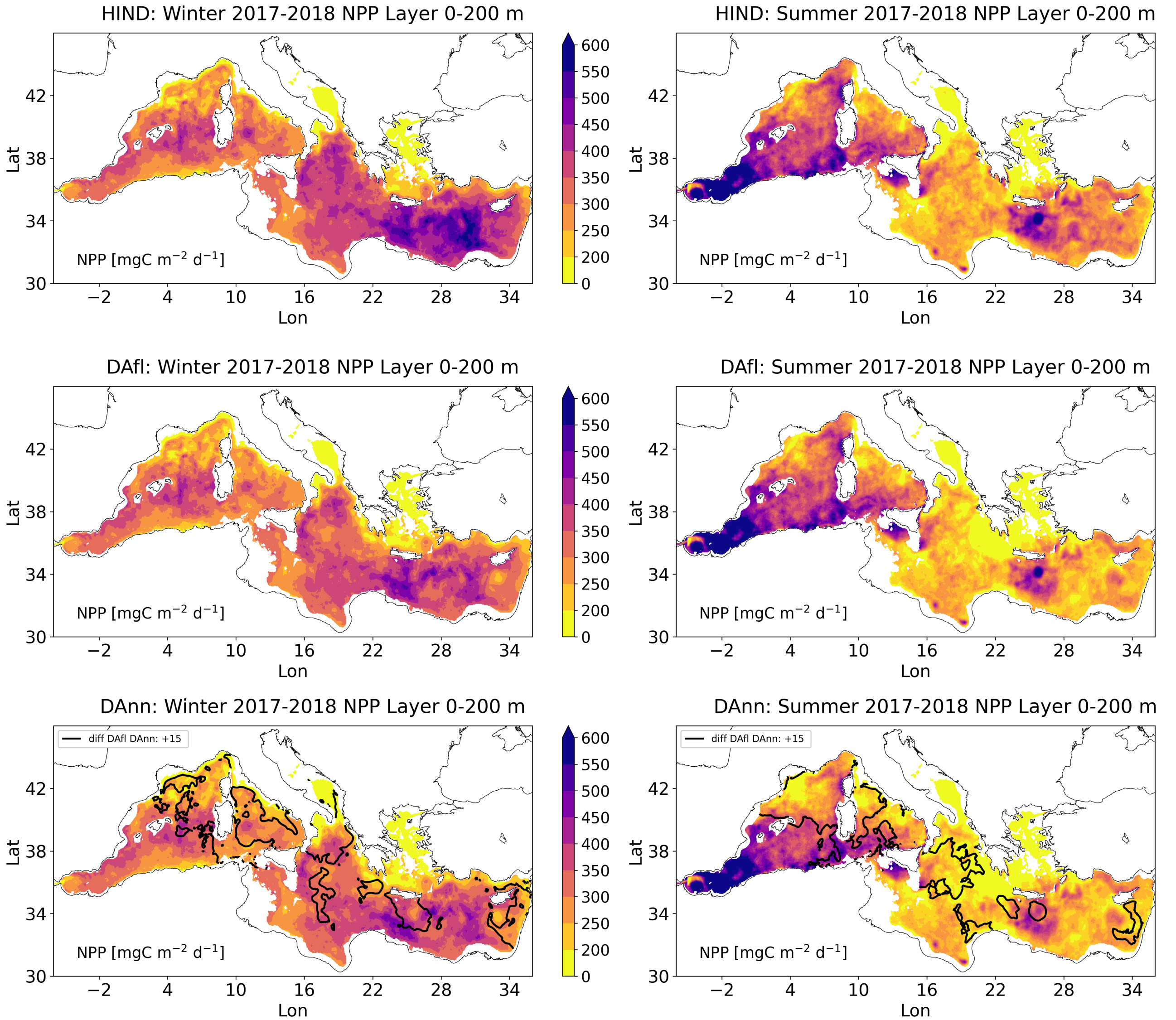

Figure 11. Maps of ~~net~~ winter (FMA) and summer (JJA) net primary production [NPP, mgC m-2 d-1] in the three ~~runs~~simulations: ~~A.~~ HIND ~~B.~~ (first row); DAfl ~~C~~(second row). DAnn (third row). Seasonal averages ~~are computed considering~~ were calculated for the period 2017-2018. The black contour lines in the third row encompass areas where the NPP difference between DAnn and DAfl exceeds 15 mgC m^{-2} d^{-1}

3.3.3 Impact on Argo Observing system design

Analyzing the departure of an assimilated simulation from a reference solution provides insights into the impact of the observing system design and several data impact indicators can be used (Ford 2021, Teruzzi et al. 2021 and Raicich and Rampazzo 2003). In this work, we ~~adopt~~ adopted the impact indicator $I_{ij}(t)$ as described in Teruzzi et al. (2021)~~, in order to quantify the~~

~~integrated (0-300m) response of~~ . This indicator supports the quantification of the vertically integrated response resulting from assimilating BGC Argo profiles ~~with respect to the no assimilative~~ compared to the non-assimilative run:

$$I_{ij}(t) = \cancel{\frac{|Sim_{ij}(t) - HIND_{ij}(t)|_{300}}{(HIND_{300})_{mean}}} \frac{|Sim_{ij}(t) - HIND_{ij}(t)|_{0-maxdepth}}{(HIND_{(0-maxdepth)})_{mean}} \tag{2}$$

~~Here, HIND is here~~ HIND is the reference, while Sim refers to one of the different DA set-ups. $|Sim_{ij}(t) - HIND_{ij}(t)|$ is the absolute difference between two simulations (for each day and grid point), while the subscript ~~300 represents the integral over the 0–300m~~ *maxdepth* indicates the vertical integrated layer of 0-300 m and 0-600 m for chlorophyll and nitrate respectively. The indicator $I_{ij}(t)$ quantifies how much an assimilated run deviates from the reference simulation (HIND) ~~simulation.~~ for every grid point of the Mediterranean Sea domain and its 95^{th} percentile permits to highlight the highest impacts.

Figures 12 and 13 show the nitrate and chlorophyll $I_{ij}(t)$ ~~95th~~ 95^{th} percentile of the seasonal indicator in winter (left column) and in summer (right column) in the DAfl (first row) and DAnn (second row) simulations. ~~The 95th percentile of the indicator shows that the BGC-Argo assimilation impacts (DAfl) both chlorophyll and nitrate.~~

In DAfl, the extent of nitrate $I_{ij}(t)$ ~~95th~~ 95^{th} above 0.1 (the mean of the 95^{th} percentile impact indicator in the Mediterranean Sea) is 16.5% and 18.7% in winter and in summer respectively, with clear spatial distribution mapping the BGC-Argo density. The introduction of reconstructed profiles in DAnn make it possible to increase the nitrate impacted areas up to about 35% and 39% in winter and summer respectively. The DAnn impact increase is mainly localized in the western Mediterranean Seas and in the ~~Ionian Sea~~Ion, while the less evident impact in the ~~Levantine~~Lev, especially in summer, is mainly due to the low number of NN reconstructed nitrate in the area.

Chlorophyll impact maps (Figure 13) show that besides the direct impact of chlorophyll profiles assimilation, phytoplankton is also affected by the reconstructed nitrate assimilation. Compared to a threshold of 0.4 (the mean of the 95^{th} percentile impact indicator in the Mediterranean Sea), the impacted areas increase from 18.2% to 29.8% in winter and from 10.8% to 14.5% in summer in the DAfl and DAnn runs. These results suggest that the inclusion of reconstructed nitrate assimilation can potentially extend the impact to almost all the Mediterranean Sea~~, with the only exclusion of~~ . The scarcity or absence of available data for assimilation prevents us from observing an impact in the marginal seas (~~Adriatic and Aegean)and~~ Adr and Aeg), the southern part of Ionian (ion1) and Western sub-basins (alb and swm1).

Oxygen impact maps (not shown) are very similar to the nitrate DAnn maps and do not show significant differences between the two DA simulation, since the same QC oxygen dataset was assimilated in DAfl and DAnn. Moreover, as detailed in Section 2.2, the oxygen assimilation exclusively updates the oxygen itself and have minimal impact on its dynamics from the other biogeochemical cycles.

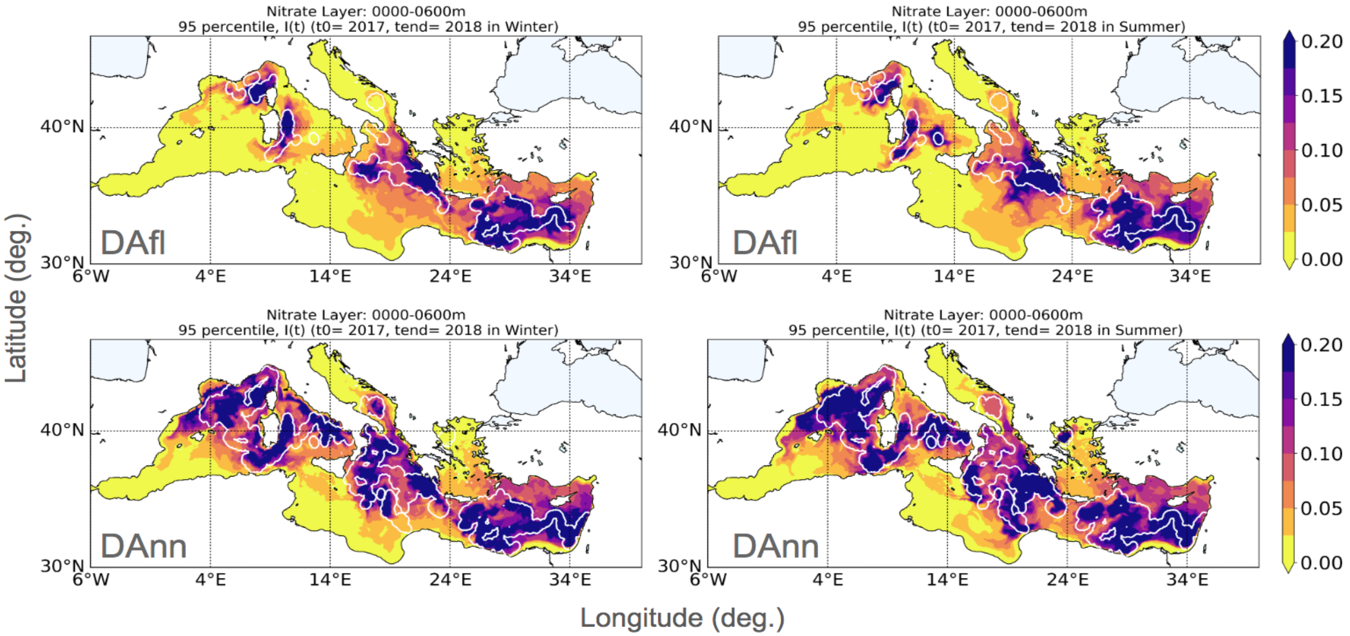

Figure 12. Maps of Iij(t) ~~95th~~ 95^{th} percentiles for Nitrate in winter (left column) and summer (right column) in the DAfl (first row) and DAnn (second row); white ~~contours~~ contour lines identify the areas within three correlation radii from the float profiles

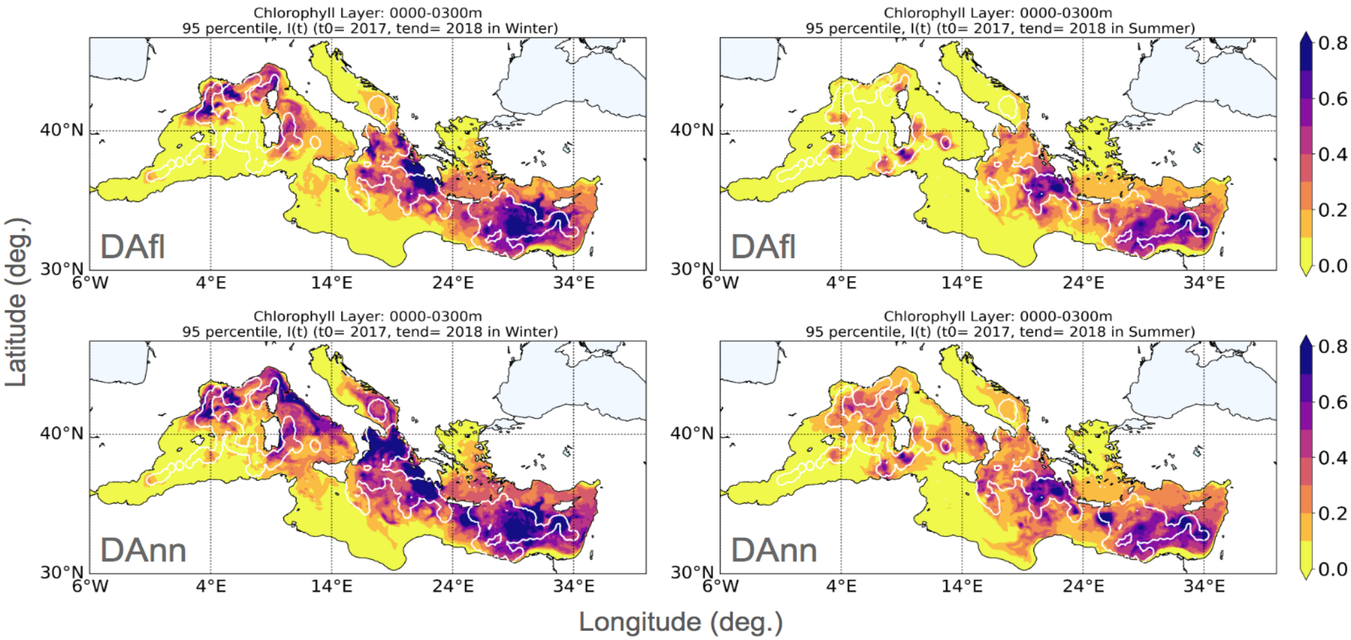

Figure 13. Maps of Iij(t) ~~95th~~ 95^{th} percentiles for Chlorophyll in winter (left column) and summer (right column) in the DAfl (first row) and DAnn (second row); white ~~contours~~ contour lines identify the areas within three correlation radii from the float profiles

4 Discussion

Our quality check procedure (QC O2) for oxygen drift detection and comparison with a reference dataset ~~can~~ successfully integrates the official ~~QC BGC-Argo, making oxygen BGC-Argo~~ BGC Argo (Argo, 2022), making the oxygen BGC Argo a robust and valuable dataset (Amadio et al., 2023) for initial conditions, data assimilation, validation and new product reconstruction. Even if the distinction between real oxygen depletion signals and optode drift can remain problematic without ~~in-situ~~ in situ high quality data, we believe that literature and prior knowledge can be used as a baseline for drift discrimination.

In particular, ~~recent studies have revealed a decrease of oxygen concentrations in the Mediterranean Sea (Sisma-Ventura et al. 2021 and Di Biagio et al. 2022); such tendency has been defined as a multidecadal shifts (Coppola et al. 2018 and Mavropoulou et al. 2020) and also patchy deoxygenation (Mancini et al., 2023). One~~ the oxygen concentration in the mesopelagic layer of the Mediterranean Sea can show basin scale (Mavropoulou et al., 2020) and local (Sisma-Ventura et al., 2021) intense multiyear variability. For example, one of the most evident signals ~~is~~ was the early 1990~~'~~'s East Mediterranean Transient (EMT) associated with the variations of thermohaline circulation ~~, which has caused a strong interannual variability of oxygen (Sisma-Ventura et al., 2021). Based on this literature and considering the recent years, a~~ which caused a negative and positive variation (e.g., about 10 $\mathrm{mmol\ m^{-3}}$ on a decadal time scale) of oxygen in the Western and Eastern Mediterranean Sea (Mavropoulou et al., 2020). However, in the last decades, a much smaller inter-annual variability of oxygen in the mesopelagic layer was observed in both western and eastern basins (Coppola et al. 2018 and Mavropoulou et al. 2020). Therefore, the threshold of $1\ \mathrm{mmol\ m^{-3}\ y}$ at 600 and 800 meters appears a prudent limit for sensor drift discrimination from real long term signals for our specific application.

Up to now, ~~the oceanographer visual check is required to distinguishing~~ oceanographer visual checks have been necessary to distinguish ocean signals from sensor drift (Wang et al., 2020) and the ~~debate on how to replace visual check to~~ ongoing debate about replacing visual checks with automatic statistical procedures is still open. ~~Thus~~Consequently, our work ~~can~~ seeks to contribute by proposing a new tool designed to automatically handle deep ocean signal or optode drift issues.

~~The search of drift is based on a robust combination of several factors: (i) the float history has to be longer than 1 year, (ii) the calculation is done with two robust non-parametric trend methods and (iii) at two different depths (i.e. 600m and 800m) sampling different water masses that can have different long-term dynamics. Lastly, the sensor drift signal must be confirmed by the four results and has to be compared with typical basin-wide trend values reported by scientific literature (i.e., the choice of the threshold).~~

~~The~~ This method can be further developed by applying ~~the~~ oxygen drift analysis at ~~some fixed isopycnalstogether with the one~~ fixed isopycnals, in conjunction with analysis at constant isobaths. ~~Thus, possible~~ This approach might allow us to filter out potential oxygen concentration changes ~~due to~~ caused by floats moving across different water masses~~can be filtered out~~.

The assimilation of vertical profiles provides complementary information with respect to satellite ocean ~~color assimilation (Cossarini et al. 2019 and Verdy and Mazloff 2017~~ colour assimilation (Verdy and Mazloff 2017 and Cossarini et al. 2019), which ~~however is still~~ remains the most commonly used in operational systems (Fennel et al., 2019). In fact, the effectiveness of the profiles assimilation, which has the capability to constrain vertical biogeochemical dynamics in subsurface layers (Kaufman et al. 2018, Teruzzi et al. 2021, Ford 2021, Skakala et al. 2021 and Wang et al. 2022), lies in the amount of available BGC-Argo data, that are generally insufficient to constrain a basin ~~wide simulation. In Teruzzi et al. (2021), results of the impact indicator principally showed the~~ wide-simulation. Previous findings (Teruzzi et al., 2021) have primarily demonstrated the efficiency of ocean ~~color~~ colour assimilation in constraining chlorophyll dynamics~~mostly during winter . In this work, important impacts are also observed in summerfor all variables, as a consequence of the increased number of assimilated~~ , especially during winter and the advantages of assimilating BGC-Argo profiles in summer. Our work highlights the larger and more extensive benefits of profile assimilation during summer due to the incorporation of reconstructed nitrate profiles.

~~In fact, through~~ Through the integration of NN and DA, the ~~number~~ count of nitrate profiles ingested can ~~significantly increase, with a density of more than 30 profiles every ten days over a basin of 2.5M . Indeed, as shown in the results, Argo and BGC-Argo oxygen sensors can potentially support~~ potentially be as high as the BGC Argo equipped with an oxygen sensor (i.e., more than double of the nitrate profiles), which corresponds to a density of ~~biogeochemical profiles up to~~ 1 profile in each 2.5~~°~~ deg x 2.5~~°~~ deg box every 10 days for the 2017-2018 period. This means that seasonal sub-basins scale dynamics (e.g., bloom or stratification) can effectively be constrained~~while, up to now,~~ , while the mesoscale dynamics can ~~probably be exclusively locally studied (d'Ortenzio et al., 2021).~~ be only locally constrained (d'Ortenzio et al., 2021).

Apart from an increase in the numbers of floats, ~~improvements in the simulation of mesoscale dynamics~~ a further increase of the area impacted from a float assimilation can be achieved by redefining horizontal covariance errors in the data assimilation scheme. Indeed, benefits of non-uniform correlation radius in the horizontal scale have been previously investigated (Cossarini et al., 2019) and additional improvements could be provided by a 3D varying correlation radius (Storto et al., 2014).

Looking at the recent evolution in the availability of BGC-Argo sensors (Figure 14), our combined NN and DA approach would allow keeping the benefits of the BGC-Argo ~~OS~~ Observing System in the Mediterranean operational system. Even if nitrate and chlorophyll profiles have dramatically decreased after 2020, the assimilation of reconstructed profiles can potentially overcome this lack. Nevertheless, as shown in our ~~OSE experiments~~ Observing System Experiment (Figure 12 and 13), there are still ~~undersampled~~ under-sampled areas by the Argo and oxygen sensors, such as Alboran, Southern Ionian seas and the marginal seas (Northern Adriatic and Northern Aegean Sea) which would require specific deployments.

With respect to previous BGC ~~OSSE experiments~~ Observing System Simulation Experiments, (Yu et al. 2018, Ford 2021), we show how to exploit the current Argo and BGC-Argo networks for reconstructing biogeochemical variables.

MLP feed-forward methods to reconstruct biogeochemical variables are good enough (Bittig et al. 2018c, Sauzède et al. 2020 Fourrier et al. 2021 and Pietropolli et al. 2023) to reach our purposes, even if their application to generate smooth and consistent profiles still has some limitations (Pietropolli et al., 2023). The MLP-NN-MED method ~~has~~ exhibits a validation error of 0.50 $\mathrm{mmol\,m^{-3}}$ for nitrate and 0.87 $\mathrm{mmol\,m^{-3}}$ when ~~it is used~~ applied to predict BGC-Argo data (Pietropolli et al., 2023). These ~~values are slightly higher than the BGC-Argo error estimated from the triple collocation method (Mignot et al., 2019), which is used as the observation error. We recognized that~~ uncertainties related to the reconstructed nitrate dataset are higher then the one used in our study (0.25 $\mathrm{mmol\,m^{-3}}$) for both BGC Argo and reconstructed profiles.

Using the same error for both datasets revealed the highest potential impact of the reconstructed nitrate. On the other hand, using a possibly underestimated error could unbalance the assimilation results toward observation over-fitting, and we recognize the potential benefits of using different error values ~~can be used~~ for BGC-Argo and reconstructed profiles~~to take account of their uncertainties. Then, it is intuitive that with higher error , the reconstructed dataset impact would have been lower~~. Over-fitting effects towards observations may similarly derive from our choice of not explicitly including the nitrate representation error. However, our nitrate error definition is an evolution of the approach used in Teruzzi et al. (2021), which demonstrated a well-established balance between assimilation impacts and over-fitting towards the observations.

~~Indeed, the~~ The larger error in MLP-NN-MED prediction of BGC-Argo profiles derives the fact that the MLP methods, being pointwise based, are unaware of the vertical gradient (e.g., typical shape) of the profiles of biogeochemical variables that they seek to infer. This fact can lead to irregularities and lack of smoothness in the predicted profiles (Pietropolli et al., 2023), which we partly solved by adding a smoothing operator. However, one way to increase the reliability of profile reconstruction would be to include information with a physical meaning from observed data (Buizza et al., 2022). 1D Convolutional Neural Networks represent a viable alternative approach considering their ability to treat the coherence of the 1D signals (e.g., typical shapes of profiles) as shown in Li et al. (2021).

Integration of NN and DA have been tested in several geoscience applications (Buizza et al. 2022, Brajard et al. 2021, Stanev et al. 2022) to infer unresolved spatial scales or reproduce missing data. In our application, the integration of NN, which retrieves a large number of profiles (Pietropolli et al., 2023), and DA, which can apply the correction to all nutrients through error covariances (Teruzzi et al., 2021), allows spatial and multivariate changes to be captured both at the local scale and across the basin to constrain Mediterranean productivity (Figure 11). Although the corrections take time to extend to the entire basin (Figure 7), our simulations have shown that constraining bottom-up ecosystem processes (e.g., productivity, organic matter sink) has proven effective and ~~should~~ might be used in conjunction with the classical ocean ~~color~~ colour correction to phytoplankton biomass.

Any plan to learn directly from observations will have to face with some challenges, such as the use of observations whose time and space coverage is uneven or related to specific processes (Geer, 2021). The modular approach followed in this work represents a successful example of exploiting the strengths of neural networks and data assimilation to enhance the observing system impact in the operational biogeochemical system of the Mediterranean Sea.

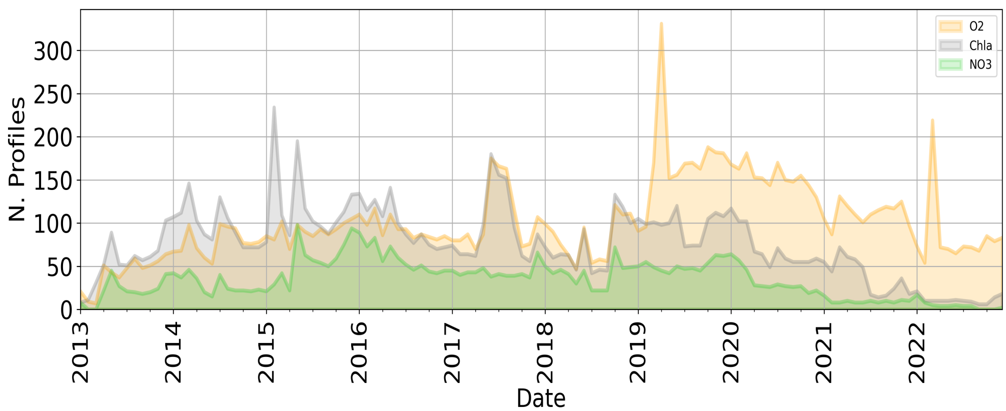

Figure 14. ~~Timeserie~~ Monthly availability of BGC-Argo profiles ~~availability~~ (~~2013-2022~~number of profiles/month) from 2013 to 2022 for ~~:~~ nitrate (green), chlorophyll (grey) and oxygen (yellow)

5 Conclusions

Combining deterministic Feed-Forward Neural Network and Data Assimilation to design an Observing System Experiment has enabled demonstrating the enhanced positive impact of profiles assimilation in the Copernicus Operational System for Short-Term Forecasting of the Biogeochemistry of the Mediterranean Sea (MedBFM).

The development of the oxygen QC procedure allowed to statistically deal with optode ~~in-situ~~ in situ drift and to derive accurate reconstructed profiles of nitrate, keeping the number of assimilated observations at a much higher level despite the current negative trend in BGC-Argo availability.

Achieved BGC profiles density provides valuable and additional information to complement that of ocean colour in describing phytoplankton seasonal blooms and stratification dynamics at sub-basins scale.

The assimilation of BGC-Argo nitrate corrects a general positive bias of the model in several Mediterranean areas, and the addition of reconstructed profiles makes the correction stronger.

Together with nitrate assimilation, the phosphate update through error covariances, sustains spatial and multivariate changes that are capable of correcting key biogeochemical processes (e.g., nitracline and deep chlorophyll maximum) and to constrain ecosystem processes (e.g., productivity) at basin-wide scale.

Author contributions. CA, AT and GCoss conceived the study. CA and AT updated the 3DVarBio code and GP and LM developed the MLP-NN-MED model. CA and GCoi performed the simulations. CA, AT and GCoss conducted the analysis of the simulation results. CA, AT, GP and GCoss wrote the draft. All authors have approved the manuscript and agree with its submission.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. This research has been partly supported by the MED-MFC "Mediterranean Monitoring and Forecasting Centre" of Copernicus Marine Service, which is implemented by Mercator Ocean International within the framework of a delegation agreement with the European Union. (Ref. n. 21002L5-COP-MFC MED-5500).

References

Amadio, C., TERUZZI, A., Feudale, L., BOLZON, G., DI BIAGIO, V., Lazzari, P., Álvarez, E., Coidessa, G., Salon, S., and COSSARINI, G.: Mediterranean Quality checked BGC-Argo 2013-2022 dataset, https://doi.org/10.5281/zenodo.10391759, 2023.

Argo: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) 95141, https://doi.org/10.17882/42182, 2022.

Argo: Argo float data and metadata from Global Data Assembly Centre (Argo GDAC), https://doi.org/10.17882/42182, 2023.

Barbieux, M., Uitz, J., Gentili, B., Pasqueron de Fommervault, O., Mignot, A., Poteau, A., Schmechtig, C., Taillandier, V., Leymarie, E., Penkerc'h, C., et al.: Bio-optical characterization of subsurface chlorophyll maxima in the Mediterranean Sea from a Biogeochemical-Argo float database, Biogeosciences, 16, 1321–1342, 2019.

Barbieux, M., Uitz, J., Mignot, A., Roesler, C., Claustre, H., Gentili, B., Taillandier, V., d'Ortenzio, F., Loisel, H., Poteau, A., et al.: Biological production in two contrasted regions of the Mediterranean Sea during the oligotrophic period: an estimate based on the diel cycle of optical properties measured by BioGeoChemical-Argo profiling floats, Biogeosciences, 19, 1165–1194, 2022.

Bittig, H. C., Körtzinger, A., Neill, C., Van Ooijen, E., Plant, J. N., Hahn, J., Johnson, K. S., Yang, B., and Emerson, S. R.: Oxygen optode sensors: principle, characterization, calibration, and application in the ocean, Frontiers in Marine Science, 4, 429, 2018a.

Bittig, H. C., Körtzinger, A., Neill, C., Van Ooijen, E., Plant, J. N., Hahn, J., Johnson, K. S., Yang, B., and Emerson, S. R.: Oxygen optode sensors: principle, characterization, calibration, and application in the ocean, Frontiers in Marine Science, 4, 429, 2018b.

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, Frontiers in Marine Science, 5, 328, 2018c.

Brajard, J., Carrassi, A., Bocquet, M., and Bertino, L.: Combining data assimilation and machine learning to infer unresolved scale parametrization, Philosophical Transactions of the Royal Society A, 379, 20200 086, 2021.

Buga, L., Sarbu, G., Fryberg, L., Magnus, W., Wesslander, K., Gatti, J., Leroy, D., Iona, S., Larsen, M., Koefoed Rømer, J., et al.: EMODnet Thematic Lot n○ 4/SI2. 749773 EMODnet Chemistry Eutrophication and Acidity aggregated datasets v2018, EMODnet [data set], 2018.

Buizza, C., Casas, C. Q., Nadler, P., Mack, J., Marrone, S., Titus, Z., Le Cornec, C., Heylen, E., Dur, T., Ruiz, L. B., et al.: Data learning: integrating data assimilation and machine learning, Journal of Computational Science, 58, 101 525, 2022.

Bushinsky, S. M., Emerson, S. R., Riser, S. C., and Swift, D. D.: Accurate oxygen measurements on modified A rgo floats using in situ air calibrations, Limnology and Oceanography: Methods, 14, 491–505, 2016.

Canu, D. M., Ghermandi, A., Nunes, P. A., Lazzari, P., Cossarini, G., and Solidoro, C.: Estimating the value of carbon sequestration ecosystem services in the Mediterranean Sea: An ecological economics approach, Global Environmental Change, 32, 87–95, 2015.

Capet, A., Stanev, E. V., Beckers, J.-M., Murray, J. W., and Grégoire, M.: Decline of the Black Sea oxygen inventory, Biogeosciences, 13, 1287–1297, 2016.

Coppini, G., Clementi, E., Cossarini, G., Salon, S., Korres, G., Ravdas, M., Lecci, R., Pistoia, J., Goglio, A. C., Drudi, M., et al.: The Mediterranean forecasting system. Part I: evolution and performance, EGUsphere, pp. 1–50, 2023.

Coppola, L., Legendre, L., Lefevre, D., Prieur, L., Taillandier, V., and Riquier, E. D.: Seasonal and inter-annual variations of dissolved oxygen in the northwestern Mediterranean Sea (DYFAMED site), Progress in Oceanography, 162, 187–201, 2018.

Cossarini, G., Lazzari, P., and Solidoro, C.: Spatiotemporal variability of alkalinity in the Mediterranean Sea, Biogeosciences, 12, 1647–1658, 2015a.

Cossarini, G., Querin, S., and Solidoro, C.: The continental shelf carbon pump in the northern Adriatic Sea (Mediterranean Sea): Influence of wintertime variability, Ecological Modelling, 314, 118–134, 2015b.

Cossarini, G., Mariotti, L., Feudale, L., Mignot, A., Salon, S., Taillandier, V., Teruzzi, A., and d'Ortenzio, F.: Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea, Ocean Modelling, 133, 112–128, 2019.

Cossarini, G., Feudale, L., Teruzzi, A., Bolzon, G., Coidessa, G., Solidoro, C., Di Biagio, V., Amadio, C., Lazzari, P., Brosich, A., et al.: High-Resolution Reanalysis of the Mediterranean Sea Biogeochemistry (1999–2019), Frontiers in Marine Science, 8, 741 486, 2021.

Dall'Olmo, G. and Mork, K. A.: Carbon export by small particles in the Norwegian Sea, Geophysical Research Letters, 41, 2921–2927, 2014.

Dang, X., Peng, H., Wang, X., and Zhang, H.: Theil-sen estimators in a multiple linear regression model, Olemiss Edu, 2008.

Di Biagio, V., Salon, S., Feudale, L., and Cossarini, G.: Subsurface oxygen maximum in oligotrophic marine ecosystems: mapping the interaction between physical and biogeochemical processes, Biogeosciences Discussions, pp. 1–33, 2022.

Dobricic, S., Pinardi, N., Adani, M., Tonani, M., Fratianni, C., Bonazzi, A., and Fernandez, V.: Daily oceanographic analyses by the Mediterranean basin scale assimilation system, Ocean Science Discussions, 3, 1977–1998, 2006.

d'Ortenzio, F., Taillandier, V., Claustre, H., Coppola, L., Conan, P., Dumas, F., Durrieu du Madron, X., Fourrier, M., Gogou, A., Karageorgis, A., et al.: BGC-Argo floats observe nitrate injection and spring phytoplankton increase in the surface layer of Levantine Sea (Eastern Mediterranean), Geophysical Research Letters, 48, e2020GL091 649, 2021.

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, Frontiers in Marine Science, 7, 120, 2020.

Escudier, R., Clementi, E., Cipollone, A., Pistoia, J., Drudi, M., Grandi, A., Lyubartsev, V., Lecci, R., Aydogdu, A., Delrosso, D., et al.: A high resolution reanalysis for the Mediterranean Sea, Frontiers in Earth Science, p. 1060, 2021.

Fennel, K., Gehlen, M., Brasseur, P., Brown, C. W., Ciavatta, S., Cossarini, G., Crise, A., Edwards, C. A., Ford, D., Friedrichs, M. A., et al.: Advancing marine biogeochemical and ecosystem reanalyses and forecasts as tools for monitoring and managing ecosystem health, Frontiers in Marine Science, 6, 89, 2019.

Feudale, L., Bolzon, G., Lazzari, P., Salon, S., Teruzzi, A., Di Biagio, V., Coidessa, G., Alvarez Suarez, E., Amadio, C., Cossarini, G., et al.: Mediterranean Sea Biogeochemical Analysis and Forecast (Copernicus Marine Service MED-Biogeochemistry, MedBFM4 system), 2022.

Fischler, M. A. and Bolles, R. C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, 24, 381–395, 1981.

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

Foujols, M.-A., Lévy, M., Aumont, O., and Madec, G.: OPA 8.1 Tracer model reference manual, Institut Pierre Simon Laplace, p. 39, 2000.

Fourrier, M., Coppola, L., Claustre, H., D'Ortenzio, F., Sauzède, R., and Gattuso, J.-P.: A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean Sea: CANYON-MED, Frontiers in Marine Science, 7, 620, 2020.

Fourrier, M., Coppola, L., Claustre, H., D'Ortenzio, F., Sauzède, R., and Gattuso, J.-P.: Corrigendum: A regional neural network approach to estimate water-column nutrient concentrations and carbonate system variables in the Mediterranean Sea: CANYON-MED, Frontiers in Marine Science, 8, 650 509, 2021.

Garcia, H., Boyer, T., Baranova, O., Locarnini, R., Mishonov, A., Grodsky, A. e., Paver, C., Weathers, K., Smolyar, I., Reagan, J., et al.: World ocean atlas 2018: Product documentation, A. Mishonov, Technical Editor, 2019.

Gasparin, F., Guinehut, S., Mao, C., Mirouze, I., Rémy, E., King, R. R., Hamon, M., Reid, R., Storto, A., Le Traon, P.-Y., et al.: Requirements for an integrated in situ Atlantic Ocean observing system from coordinated observing system simulation experiments, Frontiers in Marine Science, 6, 83, 2019.

Geer, A.: Learning earth system models from observations: machine learning or data assimilation?, Philosophical Transactions of the Royal Society A, 379, 20200 089, 2021.

Hollingsworth, A., Shaw, D., Lönnberg, P., Illari, L., Arpe, K., and Simmons, A.: Monitoring of observation and analysis quality by a data assimilation system, Monthly Weather Review, 114, 861–879, 1986.

Hornik, K., Stinchcombe, M., and White, H.: Multilayer feedforward networks are universal approximators, Neural networks, 2, 359–366, 1989.

Johnson, K. and Claustre, H.: Bringing biogeochemistry into the Argo age, Eos, Transactions American Geophysical Union, 2016.

Johnson, K., Maurer, T., Plant, J., Bittig, H., Schallenberg, C., and Schmechtig, C.: BGC-Argo quality control manual for nitrate concentration, 2021.

Johnson, K. S., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Swift, D. D., and Riser, S. C.: Long-term nitrate measurements in the ocean using the In Situ Ultraviolet Spectrophotometer: sensor integration into the Apex profiling float, Journal of Atmospheric and Oceanic Technology, 30, 1854–1866, 2013.

Johnson, K. S., Plant, J. N., Riser, S. C., and Gilbert, D.: Air oxygen calibration of oxygen optodes on a profiling float array, Journal of Atmospheric and Oceanic Technology, 32, 2160–2172, 2015.

Kaufman, D. E., Friedrichs, M. A., Hemmings, J. C., and Smith Jr, W. O.: Assimilating bio-optical glider data during a phytoplankton bloom in the southern Ross Sea, Biogeosciences, 15, 73–90, 2018.

Kumar, S., Peters-Lidard, C., Santanello, J., Reichle, R., Draper, C., Koster, R., Nearing, G., and Jasinski, M.: Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes, Hydrology and Earth System Sciences, 19, 4463–4478, 2015.

Lary, D. J., Zewdie, G. K., Liu, X., Wu, D., Levetin, E., Allee, R. J., Malakar, N., Walker, A., Mussa, H., Mannino, A., et al.: Machine learning applications for earth observation, Earth observation open science and innovation, 165, 2018.

Lazzari, P., 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, 2012.

Lazzari, P., Solidoro, C., Salon, S., and Bolzon, G.: Spatial variability of phosphate and nitrate in the Mediterranean Sea: A modeling approach, Deep Sea Research Part I: Oceanographic Research Papers, 108, 39–52, 2016.

Le Traon, P.-Y.: From satellite altimetry to Argo and operational oceanography: three revolutions in oceanography, Ocean Science, 9, 901–915, 2013.

Le Traon, P. Y., Reppucci, A., Alvarez Fanjul, E., Aouf, L., Behrens, A., Belmonte, M., Bentamy, A., Bertino, L., Brando, V. E., Kreiner, M. B., et al.: From observation to information and users: The Copernicus Marine Service perspective, Frontiers in Marine Science, 6, 234, 2019.

Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J.: A survey of convolutional neural networks: analysis, applications, and prospects, IEEE transactions on neural networks and learning systems, 2021.

Li, Z. Q., Liu, Z. H., and Lu, S. L.: Global Argo data fast receiving and post-quality-control system, IOP Conference Series: Earth and Environmental Science, 502, 012 012, https://doi.org/10.1088/1755-1315/502/1/012012, 2020.

Mancini, A. M., Bocci, G., Morigi, C., Gennari, R., Lozar, F., and Negri, A.: Past Analogues of Deoxygenation Events in the Mediterranean Sea: A Tool to Constrain Future Impacts, Journal of Marine Science and Engineering, 11, 562, 2023.

Marañón, E., Van Wambeke, F., Uitz, J., Boss, E. S., Dimier, C., Dinasquet, J., Engel, A., Haëntjens, N., Pérez-Lorenzo, M., Taillandier, V., et al.: Deep maxima of phytoplankton biomass, primary production and bacterial production in the Mediterranean Sea, Biogeosciences, 18, 1749–1767, 2021.

Martinez, E., Brini, A., Gorgues, T., Drumetz, L., Roussillon, J., Tandeo, P., Maze, G., and Fablet, R.: Neural network approaches to reconstruct phytoplankton time-series in the global ocean, Remote Sensing, 12, 4156, 2020a.

Martinez, E., Gorgues, T., Lengaigne, M., Fontana, C., Sauzède, R., Menkes, C., Uitz, J., Di Lorenzo, E., and Fablet, R.: Reconstructing global chlorophyll-a variations using a non-linear statistical approach, Frontiers in Marine Science, 7, 464, 2020b.

Maurer, T. L., Plant, J. N., and Johnson, K. S.: Delayed-mode quality control of oxygen, nitrate, and pH data on SOCCOM biogeochemical profiling floats, Frontiers in Marine Science, 8, 683 207, 2021.

Mavropoulou, A.-M., Vervatis, V., and Sofianos, S.: Dissolved oxygen variability in the Mediterranean Sea, Journal of Marine Systems, 208, 103 348, 2020.

Mignot, A., 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, Global Biogeochemical Cycles, 28, 856–876, 2014.

Mignot, A., d'Ortenzio, F., Taillandier, V., Cossarini, G., and Salon, S.: Quantifying observational errors in Biogeochemical-Argo oxygen, nitrate, and chlorophyll a concentrations, Geophysical Research Letters, 46, 4330–4337, 2019.

Miloslavich, P., Seeyave, S., Muller-Karger, F., Bax, N., Ali, E., Delgado, C., Evers-King, H., Loveday, B., Lutz, V., Newton, J., et al.: Challenges for global ocean observation: the need for increased human capacity, Journal of Operational Oceanography, 12, S137–S156, 2019.

Oddo, P., Adani, M., Pinardi, N., Fratianni, C., Tonani, M., and Pettenuzzo, D.: A nested Atlantic-Mediterranean Sea general circulation model for operational forecasting, Ocean science, 5, 461–473, 2009.

Pietropolli, G., Manzoni, L., and Cossarini, G.: Multivariate Relationship in Big Data Collection of Ocean Observing System, Applied Sciences, 13, 5634, 2023.

Pinardi, N., Zavatarelli, M., Adani, M., Coppini, G., Fratianni, C., Oddo, P., Simoncelli, S., Tonani, M., Lyubartsev, V., Dobricic, S., et al.: Mediterranean Sea large-scale low-frequency ocean variability and water mass formation rates from 1987 to 2007: A retrospective analysis, Progress in Oceanography, 132, 318–332, 2015.

Raicich, F. and Rampazzo, A.: Observing System Simulation Experiments for the assessment of temperature sampling strategies in the Mediterranean Sea, Annales Geophysicae, 21, 151–165, https://doi.org/10.5194/angeo-21-151-2003, 2003.

Ricour, F., Capet, A., d'Ortenzio, F., Delille, B., and Grégoire, M.: Dynamics of the deep chlorophyll maximum in the Black Sea as depicted by BGC-Argo floats, Biogeosciences, 18, 755–774, 2021.

Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., and Martinez, E.: A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers, Frontiers in Marine Science, 10, 2023.

Sakov, P. and Sandery, P.: An adaptive quality control procedure for data assimilation, Tellus A: Dynamic Meteorology and Oceanography, 69, 1318 031, 2017.

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 Science, 15, 997–1022, 2019.

Sauzède, R., Claustre, H., Uitz, J., Jamet, C., Dall'Olmo, G., d'Ortenzio, F., Gentili, B., Poteau, A., and Schmechtig, C.: A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient, Journal of Geophysical Research: Oceans, 121, 2552–2571, 2016.

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, Frontiers in Marine Science, 4, 128, 2017.

Sauzède, R., Johnson, J., Claustre, H., Camps-Valls, G., and Ruescas, A.: Estimation of oceanic particulate organic carbon with machine learning, ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2, 949–956, 2020.

Siokou-Frangou, I., Christaki, U., Mazzocchi, M. G., Montresor, M., Ribera d'Alcalá, M., Vaqué, D., and Zingone, A.: Plankton in the open Mediterranean Sea: a review, Biogeosciences, 7, 1543–1586, 2010.

Sisma-Ventura, G., Kress, N., Silverman, J., Gertner, Y., Ozer, T., Biton, E., Lazar, A., Gertman, I., Rahav, E., and Herut, B.: Post-eastern mediterranean transient oxygen decline in the deep waters of the southeast mediterranean sea supports weakening of ventilation rates, Frontiers in Marine Science, 7, 598 686, 2021.

Skakala, J., Ford, D., Bruggeman, J., Hull, T., Kaiser, J., King, R. R., Loveday, B., Palmer, M. R., Smyth, T., Williams, C. A., et al.: Towards a multi-platform assimilative system for North Sea biogeochemistry, Journal of Geophysical Research: Oceans, 126, e2020JC016 649, 2021.

Stanev, E. V., Wahle, K., and Staneva, J.: The synergy of data from profiling floats, machine learning and numerical modeling: Case of the Black Sea euphotic zone, Journal of Geophysical Research: Oceans, 127, e2021JC018 012, 2022.

Storto, A., Dobricic, S., Masina, S., and Di Pietro, P.: Assimilating along-track altimetric observations through local hydrostatic adjustment in a global ocean variational assimilation system, Monthly Weather Review, 139, 738–754, 2011.

Storto, A., Masina, S., and Dobricic, S.: Estimation and impact of nonuniform horizontal correlation length scales for global ocean physical analyses, Journal of Atmospheric and Oceanic Technology, 31, 2330–2349, 2014.

Storto, A., De Magistris, G., Falchetti, S., and Oddo, P.: A neural network–based observation operator for coupled ocean–acoustic variational data assimilation, Monthly Weather Review, 149, 1967–1985, 2021.

Taillandier, V., Prieur, L., d'Ortenzio, F., Ribera d'Alcalà, M., and Pulido-Villena, E.: Profiling float observation of thermohaline staircases in the western Mediterranean Sea and impact on nutrient fluxes, Biogeosciences, 17, 3343–3366, 2020.

Takeshita, Y., Martz, T. R., Johnson, K. S., Plant, J. N., Gilbert, D., Riser, S. C., Neill, C., and Tilbrook, B.: A climatology-based quality control procedure for profiling float oxygen data, Journal of Geophysical Research: Oceans, 118, 5640–5650, 2013.

Teruzzi, A., Dobricic, S., Solidoro, C., and Cossarini, G.: A 3-D variational assimilation scheme in coupled transport-biogeochemical models: Forecast of Mediterranean biogeochemical properties, Journal of Geophysical Research: Oceans, 119, 200–217, 2014.

Teruzzi, A., Bolzon, G., Salon, S., Lazzari, P., Solidoro, C., and Cossarini, G.: Assimilation of coastal and open sea biogeochemical data to improve phytoplankton simulation in the Mediterranean Sea, Ocean Modelling, 132, 46–60, 2018.

Teruzzi, A., Bolzon, G., Feudale, L., and Cossarini, G.: Deep chlorophyll maximum and nutricline in the Mediterranean Sea: emerging properties from a multi-platform assimilated biogeochemical model experiment, Biogeosciences, 18, 6147–6166, 2021.

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, 2019.

Thierry, V. and Bittig, H.: Argo quality control manual for dissolved oxygen concentration, Report (qualification paper (procedure, accreditation support)), FRANCE, https://doi.org/https://doi.org/10.13155/46542, 2021.

Verdy, A. and Mazloff, M. R.: A data assimilating model for estimating S outhern O cean biogeochemistry, Journal of Geophysical Research: Oceans, 122, 6968–6988, 2017.

Vichi, M., Pinardi, N., and Masina, S.: A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory, Journal of Marine Systems, 64, 89–109, 2007a.

Vichi, M., Pinardi, N., and Masina, S.: A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory, Journal of Marine Systems, 64, 89–109, 2007b.

Waller, J. A., García-Pintado, J., Mason, D. C., Dance, S. L., and Nichols, N. K.: Assessment of observation quality for data assimilation in flood models, Hydrology and Earth System Sciences, 22, 3983–3992, 2018.

Wang, B. and Fennel, K.: An assessment of vertical carbon flux parameterizations using backscatter data from BGC Argo, Geophysical Research Letters, 50, e2022GL101 220, 2023.

Wang, B., Fennel, K., Yu, L., and Gordon, C.: Assessing the value of biogeochemical Argo profiles versus ocean color observations for biogeochemical model optimization in the Gulf of Mexico, Biogeosciences, 17, 4059–4074, 2020.

Wang, B., Fennel, K., and Yu, L.: Can assimilation of satellite observations improve subsurface biological properties in a numerical model? A case study for the Gulf of Mexico, Ocean Science, 17, 1141–1156, 2021a.

Wang, S., Flipo, N., Romary, T., and Hasanyar, M.: Particle filter for high frequency oxygen data assimilation in river systems, Environmental Modelling & Software, 151, 105 382, 2022.

Wang, T., Chai, F., Xing, X., Ning, J., Jiang, W., and Riser, S. C.: Influence of multi-scale dynamics on the vertical nitrate distribution around the Kuroshio Extension: An investigation based on BGC-Argo and satellite data, Progress In Oceanography, 193, 102 543, 2021b.

Yu, L., Fennel, K., Bertino, L., El Gharamti, M., and Thompson, K. R.: Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling, Ocean Modelling, 126, 13–28, 2018.

Yumruktepe, V. Ç., Mousing, E. A., Tjiputra, J., and Samuelsen, A.: An along-track biogeochemical Argo modelling framework, a case study of model improvements for the Nordic Seas, Geoscientific Model Development Discussions, 2023, 1–31, 2023.

Zhou, H., Yue, X., Lei, Y., Tian, C., Ma, Y., and Cao, Y.: Large contributions of diffuse radiation to global gross primary productivity during 1981–2015, Global Biogeochemical Cycles, 35, e2021GB006 957, 2021.