Marine Ecosystem forecasts: skill performance of the CMEMS Mediterranean Sea model system

Stefano Salon[1], Gianpiero Cossarini[1], Giorgio Bolzon[1], Laura Feudale[1], Paolo Lazzari[1], Anna Teruzzi[1], Cosimo Solidoro[1], Alessandro Crise[1]

5 [1]Oceanography Section, OGS - Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Trieste, 34151, Italy

Correspondence to: Stefano Salon (ssalon@inogs.it)

Abstract. The quality of the upgraded version of the CMEMS biogeochemical operational system of the Mediterranean Sea (MedBFM) is assessed in terms of consistency and forecast skill, following a mixed validation protocol that exploits different reference data from satellite, oceanographic databases, Biogeochemical Argo floats, literature. We demonstrate that the

10 GODAE metrics paradigm can be efficiently applied to validate an operational model system for biogeochemical and ecosystem forecasts. The accuracy of the CMEMS biogeochemical products for Mediterranean Sea can be achieved from basin-wide and seasonal scale to mesoscale and weekly scale, and its level depends on the specific variable and the availability of reference data. In particular, the use of the Biogeochemical Argo floats data allows for a relevant enhancement of the validation framework of operational biogeochemical models, providing new skill metrics for key biogeochemical processes

15 and dynamics (e.g. deep chlorophyll maximum depth), which can be easily implemented to routinely monitor the quality of the products and highlight any possible anomaly.

1 Introduction

Operational ocean forecasting systems integrate remote observations, in situ measurements and modelling systems, and have been widely recognized as important assets for ocean state monitoring (von Schuckmann et al., 2016) and the development of

20 the blue economy (She et al., 2016). In such framework, the operational monitoring and forecasting of marine biogeochemistry and ecosystem dynamics is based on biogeochemical models designed to represent the lower trophic level ecosystem (i.e. from phytoplankton to zooplankton). The improvement of their predictive capability on weekly and seasonal time-scales mostly required by users is under development and it is strongly related to the development of data assimilation capacity, while their quality assessment is constrained to the availability of reference data, both remote and in situ (Gehlen et al., 2015) and possibly

25 independent (i.e., not assimilated; Gregg et al., 2009). In this perspective, efforts to establish a stronger link between operational biogeochemistry products and potential users from the fisheries and environmental science communities are constantly increasing (Berx et al., 2011; Payne et al., 2017).

The European Copernicus Marine Environment Monitoring Service (CMEMS; marine.copernicus.eu) operationally provides "regular and systematic core reference information" on the state, variability and dynamics of the ocean, marine ecosystems

and sea ice for the global ocean and the European regional seas (Le Traon et al., 2017). As user-driven service based on a "continuous improvement" philosophy, CMEMS is committed to maintain its operational systems up-to-date in order to supply quality-assessed products for the analysis of the current state of oceans and seas, for the short-term forecasts and for the reanalysis and reprocessing of the recent decades.

5 The CMEMS products are delivered to users through a service portfolio, where the information is organized in terms of the data origin, that is from model or from observations (satellite and in situ). Model data from analysis/forecast and reanalyses are geographically grouped for the global ocean and for six regional European regions, with a total of seven specific model systems. Each model system features a physical, a biogeochemical and a wave component.

The Mediterranean Sea Monitoring and Forecasting Centre (Med-MFC; Tonani et al., 2013) is one of the regional systems and
10 it is composed by the physical system "Med-PHY" (Tonani et al., 2008; Oddo et al. 2009), which drives the biogeochemical system "Med-BIO" (Lazzari et al., 2010, 2012) and the wave system "Med-WAV" (Günther and Behrens, 2012). In the recent years, following the CMEMS requirements, the Med-MFC has been consistently upgraded in the physical (Oddo et al. 2014; Clementi et al., 2017; Pistoia et al., 2018), wave (Zacharioudaki et al. 2018) and biogeochemical components (Cossarini et al., 2015; Lazzari et al., 2016), also including data assimilation (Teruzzi et al., 2014; Storto et al., 2015).

15 More specifically, the last major upgrade of Med-BIO focused on the increase of horizontal resolution from 1/16 to 1/24 degree. In addition, the upgrade involved different aspects of the forecasting system aimed to improve the alignment with the physical component: the new non-linear free-surface curvilinear z*-coordinate configuration used in NEMO3.6 (see Madec, 2016, for the NEMO implementation and further details) and the terrestrial input boundary conditions layout, now including 39 rivers. Moreover, Med-BIO improved the former data assimilation scheme (Teruzzi et al., 2014), extending the assimilation
20 of surface chlorophyll concentration to coastal areas (Teruzzi et al., 2018a) and reducing the time-to-solution through a parallelization of the cost function solver (Teruzzi et al., 2018b).

Both historical and near-real time data of observation systems are strategical to evaluate the quality of operational oceanography products (She et al., 2016). The assessment of the operational ocean products accuracy already benefits from international intercomparison initiatives (e.g., the GOV Task Team for Intercomparison and Validation, ICV-TT; Bell et al.,
25 2015), which also define specific protocols to quantify the quality level of core variables delivered to users (Hernandez et al., 2015; Ryan et al., 2015). Although operational ocean models are designed to span the whole water column from the surface to the bottom and are now reaching the sub-mesoscale description, deeper ocean and mesoscale remain still not adequately sampled by the operational observation systems (Bell et al, 2015; Hernandez et al., 2018).

In the present paper, we focus on the CMEMS Mediterranean Biogeochemical analysis and forecast system products delivered
30 from April 2018. According to the definition adopted within Copernicus community (Hernandez et al., 2018), the model validation follows two main tasks:

1. The pre-operational qualification is performed when a new version of the system is developed and a full range of validation metrics is applied to provide an evaluation of the skill performance of the model. The qualification is

carried out over a short reanalysis run (e.g. a couple of years) which then provides the initial conditions for the operational analysis and forecast run.

2. The routine, near-real time (NRT) validation task of forecast products is performed operationally based on the available NRT observations and it provides an evaluation of the skill performance of the analysis and forecast
5   products.

The paper is organized as follows. In Section 2 we present the MedBFM system, that is the core of the Med-BIO operational workflow, followed by the reference observations using the recently available BGC-Argo floats data (Section 3). In Section 4 the validation framework is presented, while the most relevant results of the pre-operational and the NRT quality assessment are shown in Section 5. Discussion and conclusions are drawn, respectively, in Sections 6 and 7.

2 The Mediterranean Sea Biogeochemical analysis and forecast system

2.1 MedBFMv2.1 model system

The Med-BIO analysis and forecast products are provided by the MedBFMv2.1 model system (Fig. 1), which consists of the coupled physical-biogeochemical OGSTM-BFM model and the 3DVarBio assimilation scheme. OGSTM-BFM (Lazzari et al., 2010, 2012, 2016; Cossarini et al., 2015, and references thereby) is designed with the OGSTM transport model, based on the OPA 8.1 system (Foujols et al., 2000) and a biogeochemical reactor featuring the Biogeochemical Flux Model (BFM; Vichi et al., 2007a,b), which describes the biogeochemical cycles of carbon (C) and macro-nutrients (nitrogen, N, phosphorus, P and silicon, Si) in terms of dynamical interactions among the dissolved inorganic, living organic and non-living organic compartments.

The model presently includes nine plankton functional types (PFTs): phytoplankton PFTs are diatoms, flagellates, picophytoplankton and dinoflagellates; heterotrophic PFTs consist of carnivorous and omnivorous mesozooplankton, bacteria, heterotrophic nanoflagellates and microzooplankton. The non-living compartment consists of three groups: labile, semilabile and refractory organic matter. The BFM model is also coupled to a carbonate system model (Cossarini et al., 2015, Melaku Canu et al., 2015), which consists of two prognostic state variables: alkalinity (ALK) and dissolved inorganic carbon (DIC) and provides pH, partial pressure of CO_2 (pCO2) and air-sea CO_2 flux.

3DVarBio is the variational data assimilation scheme for the correction of the four phytoplankton PFTs of BFM using surface chlorophyll retrieved from satellite observations provided by the Ocean Colour Thematic Assembly Centre (OC-TAC) of CMEMS. The 3DVarBio scheme (see details in Teruzzi et al., 2014) decomposes the background error covariance matrix using a sequence of different operators that account separately for the vertical covariance (VV), the horizontal covariance (VH) and the covariance among biogeochemical variables (Vb). VV is defined by a set of synthetic profiles that are evaluated by means of an Empirical Orthogonal Function (EOF) decomposition applied to a validated multi-annual run (over the period 1998-2015). EOFs are computed for 12 months and 30 coastal and open sea sub-regions in order to account for the variability of 3D chlorophyll anomaly fields. Surface chlorophyll is assimilated over the whole domain, including the coastal areas

(Teruzzi et al., 2018a), through the upgrade of the non-homogeneous VV and the non-uniform and direction-dependent VH specifically focussed for the case 2 waters. Further, the time-to-solution of 3DVarBio has been significantly reduced (Teruzzi et al., 2018b) using the domain decomposition with message passing paradigm to parallelize the code and maximize performance and scalability and adopting the efficient parallel solver of the PETSc/TAO library for optimizing the cost function minimization.

The MedBFMv2.1 system works on a geographical domain that spans from 9°W to 36°E and 30°N to 46°N with a meshgrid based on 1/24° longitudinal scale factor and on 1/24°cos(φ) latitudinal scale factor. The vertical meshgrid accounts for 141 vertical z-levels: 35 in the first 200 m depth, 60 between 200 and 2000 m, 28 between 2000 and 4000 m and 18 below 4000. MedBFMv2.1 features the non-linear free surface formulation (Madec et al., 2016) and includes the terrestrial inputs (e.g. nutrients, carbon and alkalinity) from 39 rivers (same as the Med-PHY) and the Dardanelles treated as a river (Fig. 2).

The MedBFM system is coupled off-line with the Med-PHY system, that provides daily 3D fields of horizontal and vertical current velocity, potential temperature, salinity, vertical eddy diffusivity, and the 2D field of sea surface height (SSH) as forcings for the OGSTM-BFM model. In particular, SSH is used in the new curvilinear z*-coordinate formulation of the MedBFM to compute the vertical scale factor which takes in account the variability of the water column volume, where the vertical coordinate follows the time-dependent non-linear variation of SSH (see Salon et al., 2018). Additional 2D fields include the surface data for solar shortwave irradiance and wind stress, which are used, respectively, to solve the gas air-sea exchanges and as input for the BFM optical module.

The Med-PHY hydrodynamics is solved by the NEMO model (v3.6; Madec et al., 2016) coupled with WaveWatch-III for the wave component and driven by atmospheric forcing of momentum, water and heat fluxes extracted by the 6-hours, 1/8 degree ECMWF operational analysis and forecast fields, plus the daily averaged precipitation and the model predicted surface temperatures (Tonani et al., 2008). The assimilation of in-situ temperature, salinity vertical profiles (VOS XBTs and ARGO floats) and along-track Sea Level Anomaly observations is performed by a variational scheme (Dobricic and Pinardi, 2008; Storto et al., 2015). Med-PHY extends into the Atlantic Ocean to accurately resolve the dynamical exchange at the Gibraltar Strait, with boundary conditions provided by the CMEMS Global analysis and forecast system products. The upgrade to the increased horizontal resolution at 1/24 degree and the validation of the CMEMS product[1] is thoroughly described in Clementi et al. (2017).

The analysis and forecast variables product available to CMEMS users for Mediterranean Sea Biogeochemistry[2] consists of 3D daily means of chlorophyll, net primary production, phytoplankton biomass, phosphate, nitrate, oxygen, pH, pCO2. The CMEMS system offers, upon free registration, the access to the 3D fields through the products catalogue and their download via ftp and https protocols (*subsetter* and *directgetfile* download).

---

[1] MEDSEA_ANALYSIS_FORECAST_PHY_006_013
[2] MEDSEA_ANALYSIS_FORECAST_BIO_006_014

2.2 Set up of the pre-operational qualification simulation for Med-BIO

The pre-operational qualification run for the Med-BIO component, carried out with MedBFMv2.1, consists of a 2-year re-analysis simulation (1 January 2016 to 31 December 2017), with set up described in the following points.

- The physical ocean (current, temperature, salinity, vertical eddy viscosity) and atmospheric (short wave radiation and wind stress) forcing daily fields are produced by the Med-PHY system and are derived from an equivalent 2-year re-analysis simulation described in Clementi et al. (2018).

- Assimilation of satellite surface chlorophyll concentration derived by the multi-sensor (MODIS and VIIRS) CMEMS product[3] of ocean colour for the Mediterranean Sea is performed by 3DVarBio.

- The initial conditions of biogeochemical variables are set as sub-basin (Fig. 2) climatological profiles computed from in-situ data collections (NODC-OGS) described in Lazzari et al. (2016) and Cossarini et al. (2015). A spin-up period of 1 year repeated for 5 times in perpetual mode is carried out before the start of the simulation.

- The biogeochemical boundary conditions are provided through a Newtonian dumping term which regulates the Atlantic buffer zone western of the Strait of Gibraltar, where the tracer concentrations are relaxed to the seasonally varying profiles. Seasonal profiles of phosphate, nitrate, silicate, dissolved oxygen are derived from an analysis of climatological MEDAR-MEDATLAS and NODC-OGS datasets, while seasonal profiles of ALK and DIC are obtained from in-situ datasets (Huertas et al., 2009; de la Paz et al., 2011; Alvarez et al., 2014).

- Nutrient (nitrogen and phosphorous) loads from 39 rivers and Dardanelles, which are aligned with the Med-PHY configuration (Fig. 2), are derived from the PERSEUS FP7-287600 project dataset (Deliverable D4.6). Using available direct observations, the nutrient discharge rates for the major rivers (Po, Rhone and Ebro) are calculated taking into account seasonal variability on a monthly scale, while the other rivers inputs are treated as constants throughout the year due to a lack of data.

- Terrestrial inputs of ALK and DIC are derived considering their typical concentrations per fresh water mass in macro coastal areas of the Mediterranean Sea and the water discharges of the 39 rivers from the PERSEUS dataset. The Dardanelles input is considered as river input: also in this case the total inflow was derived considering typical water mass concentration of ALK and DIC for Marmara Sea (Chopin-Montegut, 1993) multiplied by the net water mass fluxes.

- Atmospheric deposition rates of inorganic nitrogen and phosphorus are set according to the synthesis proposed by Ribera d'Alcalà et al. (2003) and based on measurements of field data (Loye-Pilot et al., 1990; Guerzoni et al., 1999; Herut and Krom, 1996; Cornell et al., 1995; Bergametti et al., 1992). Atmospheric deposition rates of nitrate and phosphate are assumed to be constant in time during the year, but with different values for the western (580 Kt N/yr

---

[3] OCEANCOLOUR_MED_CHL_L3_NRT_OBSERVATIONS_009_040

and 16 Kt P/yr) and eastern (558 Kt N/yr and 21 Kt P/yr) sub-basins. The rates are calculated by averaging the "low" and "high" estimates proposed by Ribera d'Alcalà et al. (2003).

- Atmospheric pCO_2 concentration is set equal to the yearly average measured at the Lampedusa station (Artuso et al., 2009) between 1992 and 2017[4] with the 2018 value extrapolated by linear regression.

- Surface evaporation-precipitation effects on dilution and concentration of tracers are directly computed by the OGSTM transport model updated with the non-linear free-surface z*-coordinate configuration.

Further details can be found in the documents available in the CMEMS catalogue (Bolzon et al., 2017).

2.3 Set up of the operational workflow for Med-BIO

The CMEMS Med-BIO operational workflow runs every Tuesday, starting after the completion of the analysis production
10 cycle of the Med-PHY workflow. The two workflows consist of 7 days of analysis (from T-7 to T-1) one day of hindcast (T0) and 10 days of forecast (from T+1 to T+10, also referred to as T1 to T10) according to the availability of the ECMWF atmospheric forcing. Additionally, in order to maintain enough number of forecast days, Med-BIO performs a new simulation of 10 forecast days on Friday, using the forecast produced by Med-PHY. Boundary conditions in the Atlantic buffer zone, rivers and atmospheric inputs are the same of the pre-operational qualification run, which constitutes the initial condition of
15 the operational system at 1 January 2018.

3 Reference datasets for validation

Surface chlorophyll data (chl_a) are derived by the multi-sensor (MODIS-AQUA and NPP-VIIRS) CMEMS product of ocean colour observations for Mediterranean Sea (see section 2.2; Volpe et al., 2007, 2012, 2017) at 1 km spatial resolution. The surface chlorophyll field combines the estimates of two algorithms for open ocean (case 1) and coastal (case 2) water types.
20 These data are usually released as NRT data within few days from the satellite overpass.

In-situ observations of chlorophyll, nitrate and oxygen concentrations are derived by the Biogeochemical-Argo (BGC-Argo) floats dataset which records start from 2013. BGC-Argo floats data are downloaded from the Argo Global Data Assembly Centre webportal and processed following the advanced product quality procedure of Schmechtig et al. (2016).

BGC-Argo chlorophyll (Chl) adjusted data are derived from real time (RT) data with a series of corrections: the quenching
25 correction (Xing et al., 2012), a re-calibration at depth (i.e., by imposing zero for Chl values below 600m), and a tuning correction (i.e., data is further divided by a factor of 2) due to a detection of an error in the manufacturer calibration of Chl fluorometer (Roesler et al., 2017). BGC-Argo nitrate concentrations (NO3) were obtained by using the Johnson and Coletti (2002) algorithm on the raw UV absorption spectrum, then corrected with quality control procedures described in Pasqueron de Fommervault et al. (2015). BGC-Argo oxygen data (O2) are estimated after the application of a quality protocol based on

---

[4] http://cdiac.ess-dive.lbl.gov/ftp/trends/co2/lampedus.co2

a linear regression constrained to pass through the origin between percent oxygen saturation values derived from the float profiles of O2 and climatological values from the World Ocean Atlas Climatology (Takeshita, 2013; Schmecthig et al., 2016). For the pre-operational period 2016-2017, the total amount of floats and profiles for each variable is given in Tab. 1.

In-situ observations of nitrate, phosphate and oxygen derived by the National Oceanographic Data Centre of OGS (NODC-OGS) dataset covering the period 1999-2013 (the list of cruises and datasets is in Lazzari et al., 2016), are used to compute reference climatological profiles for the sub-basins of Fig. 2. In-situ observations of DIC, ALK and pH (the list of cruises and dataset sources is in Cossarini et al., 2015) are used to compute reference climatological annual profiles in the sub-basins of Fig. 2. Literature data of net primary production are based on multi-annual simulation (Lazzari et al., 2012), satellite model (Colella, 2006) and in-situ estimates (Siokou-Frangou et al., 2010), here used to validate the basin-scale consistency of the corresponding model product.

4 Product quality assessment framework

The assessment of the CMEMS Mediterranean Sea biogeochemical model system is subdivided in two tasks: pre-operational qualification of the model system and routine (or NRT) validation of forecast products. The aim of the pre-operational assessment is to verify the model consistency, that is its capability of reproducing the salient characteristics of the Mediterranean Sea ecosystem comparing a short reanalysis run with historical datasets, climatology and literature estimates. The time scale of the comparison ranges from daily to seasonal. On the other hand, the operational assessment relays on the NRT observation availability and aims to evaluate the forecast skills with a temporal scale of days.

4.1 Pre-operational quality assessment

The pre-operational qualification is performed at the release of the new CMEMS version using GODAE-like metrics (see Hernandez et al., 2015 for a recent review) applied to the 2-years run pre-operational run described in Sect. 2.2. In particular, the pre-operational validation consists of the "Class 1" metrics, which quantifies the model capability to be consistent with the large-scale climatological description of the ocean processes and, for a subset of variables (i.e., chlorophyll, nitrate and oxygen), of the "Class 4" metrics, which quantify the differences between model and observations at their location and time. When chlorophyll satellite data are used, the comparison of the model and observations is evaluated before the assimilation (i.e., after 7 days of simulation w.r.t. the previous assimilation cycle) using statistics on the innovation, thus providing a forecast skill metrics (Mattern et al., 2018).

For each BGC-Argo float, the vertical profiles of chlorophyll, nitrate and oxygen are matched-up with the model results at the same position and date, producing time series of paired model and observation profiles. Considering the relevance of the seasonal evolution of the chlorophyll vertical profile in the Mediterranean Sea and the importance of analysing the vertical profile as a whole (Lavigne et al., 2015), along with classical observation-model metrics, we developed new metrics that synthesize the model capability to reproduce key elements of the vertical profile shape:

- BIAS and RMS of the difference (RMSD) between model and float of the vertically mixed winter bloom (MWB) depth, defined as the depth at which chlorophyll concentration is 10% of surface concentration during winter (from January to March);

- BIAS and RMSD between model and float of the summer deep chlorophyll maximum (DCM) depth, defined as the depth of the chlorophyll maximum below 40 m during summer (from April to October);

- BIAS and RMSD between model and float of the surface chlorophyll and nitrate concentration (SURF), and of the 0-200 m vertical average of chlorophyll and nitrate (INTG);

- correlation between each couple of chlorophyll (oxygen and nitrate) profiles from model and BGC-Argo float;

- BIAS and RMSD between model and float of the depth of the nitracline, defined as the depth (i) where the nitrate concentration is 2 mmol/m^3 (NITRCL1), and (ii) corresponding to the maximum nitrate vertical gradient (NITRCL2).

4.2 Near-real time (NRT) validation of operational forecast products

The operational skill assessment is performed at weekly frequency considering the results of the previous forecast production cycle and the NRT operational observations (satellite and BGC-Argo floats data) available within one week from the observation time. Thus, at NRT scale, simulated surface chlorophyll of the first, second, third and fourth day of forecast (i.e. forecast lead time of T0, T1, T2 and T3) is compared with the correspondent daily surface chlorophyll from satellite observations, and RMSD and BIAS between model and observations are computed and averaged over the 16 sub-basins. Moreover, all the BGC-Argo float profiles operationally available are compared with the forecast (from T0 to T6) and statistics are reported as weekly time series of RMSD and BIAS between model output of chlorophyll, nitrate and oxygen and observations. The forecast skill assessment is then compared with the results of the reference pre-operational assessments, which act as a benchmark.

5 Results

5.1 Pre-operational qualification run

5.1.1 Model consistency

To evaluate the model consistency (GODAE Class 1 metrics) with the general features of the biogeochemistry of the Mediterranean Sea in terms of chlorophyll, nutrients (nitrate and phosphate), dissolved oxygen, carbonate system variables (DIC, ALK, pCO2, pH), and primary production, model mean fields are compared with different reference datasets.

The MedBFM surface chlorophyll for the period 2016-2017 is compared with satellite data in Fig. 3, while time series of model and satellite data are shown for four selected sub-basins in Fig. 4. The Mediterranean Sea presents a high spatial heterogeneity, with sub-basins characterized by different biogeochemical dynamics (Lazzari et al., 2012). The basin-scale characteristics, widely described in literature and clearly visible in the maps of Fig. 3 and in the time series of Fig. 4, are the

higher chlorophyll concentrations and the larger seasonal cycle proper of the western sub-basins (e.g. nwm, swm2) with respect to the eastern ones (e.g. ion1, lev2). The MedBFM model correctly simulates the interannual variability observed in the difference between spring blooms in 2016 and in 2017 in swm2, with the former less intense than the latter. A slight model overestimation is observed in alb (Fig. 3), which is probably due to an overestimation of nutrient incoming fluxes at the

5   Gibraltar Strait. Finally, modelled late winter-early spring surface bloom in nwm appears anticipated of 2-3 weeks w.r.t. satellite ones.

The MedBFM nitrate and phosphate are in good agreement with the average values and shape of the climatological profiles along the Mediterranean sub-basins (Fig. 5). In particular, the model profiles are within the range of variability of the OGS-NODC climatological profiles (Fig. 5), and the correlation values are generally larger than 0.9 (Fig. 7), which corroborates the

10   very good performance of the MedBFM model in reproducing the deepening of the nutricline and the decreasing concentration values of the deep layers from the western to the eastern sub-basins.

On average, the RMSD of nitrate is 0.6 mmol/m^3 in the upper layers (0-60 m) and around 1 mmol/m^3 in the layers below; we observe a general model underestimation of about 30% of the average values at the different depths. Phosphate RMSD is below 0.03 mmol/m^3 in the 0-100 m layer, and around 0.04 mmol/m^3 in the deeper layers, while BIAS ranges between -0.03 and 0.02

15   mmol/m^3 (Fig. 7). When normalized by the standard deviation of the reference data, the surface layers show the highest uncertainty (i.e., normalised RMSD up to 1 and 1.2 for nitrate and phosphate, respectively) and a relatively low correlation. This is because the surface layers show the lowest concentration values and quite low dispersion of the values among sub-basins. Indeed, simulating nutrient concentration in the layer above the nutricline might be critical, and validation based on climatological datasets might be not fully appropriate.

20   Modelled monthly oxygen profiles result pretty well in agreement with the climatological ones and generally within the observed variability (Fig. 5; see also the very high correlation values in Fig. 7), with BIAS and RMSD lower than 11 mmol/m^3 in all selected layers (Fig. 7). The range of RMSD values correspond to 60-140% of the mean spatial basin-wide range at the different layers, but less than 15% considering the surface temporal seasonal cycle.

Figure 6 shows that the model well simulates the vertical structure of DIC and ALK, mostly within the range of variability of

25   the climatological profiles. In particular, it can be noted that the heterogeneity of the vertical profiles of DIC and ALK (i.e., the S-shape of western sub-basin profiles, specifically alb and swm2, due to the interaction of surface Atlantic waters and deep Mediterranean waters, and the almost homogeneous vertical profiles for the eastern sub-basins) is fairly well reproduced by the model. For both DIC and ALK, the mean RMSD is around 20 μmol/kg, with higher values for the upper layers. Normalized by the standard deviation of the reference data, the mean errors are 0.40 and 0.70 for ALK and DIC, respectively (Fig. 7).

30   Correlation values are both higher than 0.7 for almost all layers showing that the basin-wide gradient of carbonate system variables is well captured by the model. The uncertainty of the carbonate system variables strongly reduces at deeper depths and the modelled vertical profiles remain within the climatological variability.

Modelled pH is corroborated using both pH data measured in total scale and reported in situ conditions, and pH data calculated by CO2sys software (Lewis and Wallace, 1998) with available DIC, ALK and other regulatory information (i.e., temperature,

salinity and concentration of phosphate and silicate). Modelled pH varies across a 8-8.1 range consistently with the observed eastward and downward positive gradient. The mean error (i.e., averaged RMSD among sub-basins) is around 0.03 in the upper layers and 0.025 in layers below 100 m, which equals almost to the mean variability of data, highlighting that small scale variability of modelled pH cannot be evaluated by the present validation framework.

5   Finally, modelled pCO2 data can be only qualitatively compared with the reconstructed data using in situ DIC, ALK and the regulatory information. Along the water column sub-basin profiles, model and reconstructed data show a comparable range of variability. However, it must be noted that the model pCO2 has a large seasonal cycles at surface since the T-dependency of the solubility, while the observations display a lower variability range (Fig. 6) due to the inadequacy of the sampling throughout the seasonal cycle (Cossarini et al., 2015).

10   Net primary production (NPP) is the measure of the net uptake of carbon by phytoplankton groups (gross primary production minus fast release processes, e.g., respiration and very labile dissolved organic matter; Vichi et al., 2015). The lack of any extensive dataset of measures of primary production in Mediterranean Sea prevents the application of quantitative metrics for the assessment of the quality of this product. A qualitative assessment of the consistency of the modelled NPP with previous estimates published in scientific literature (Tab. 2) reveals that the simulated relevant gradients between eastern and western

15   regions and averaged NPP values in the different sub-basins are in good agreement with both basin-wide and sub-basin averages of previous model and satellite assessments. Estimates derived from in-situ measurements (Siokou-Frangou et al., 2010) confirm the east to west gradient simulated by the model, though the eastern values appears overestimated by MedBFM.

5.1.2 Model skill performance

Skill performance statistics based on a model *vs* observation comparison (GODAE Class 4 metrics) are computed for
20   chlorophyll, nitrate and oxygen, and represent a stricter assessment of the model performance to capture the biogeochemical temporal dynamics and mesoscale spatial variability.

First, timeseries of RMSD and BIAS of the model-satellite chlorophyll misfit are computed prior the assimilation (i.e. using satellite data that are not yet assimilated), thus representing a short-term (i.e., after 7 days from the previous assimilation cycle) skill forecast metrics (Mattern et al. 2018). Then, the mean of the BIAS and RMSD timeseries is calculated for two selected
25   seasons (i.e., from January to April, WIN, and from June to September, SUM) and reported in Fig. 8, which is completed by the mean spatial standard deviation of observations for each sub-basin.

The western sub-basins have higher uncertainty (i.e. higher RMSD) than the eastern ones, however never exceeding 0.1 mg/m³ on average, and larger during winter period because the variability of the chlorophyll is higher than during summer (Fig. 8). The relatively high values of RMSD in the western sub-basins in winter (nwm in particular) are related to the bloom dynamics,
30   which is estimated 2-3 weeks earlier by the model (Fig. 4). In these areas, blooms are strongly related to the presence of sub-mesoscale local patches, fronts, horizontal circulation structures and local mixing conditions of the water column (e.g. Macias et al., 2018) whose magnitude, timing and spatial pattern might not be completely well resolved by the sub-grid parameterizations of the Med-PHY model component that drives MedBFM, thus resulting in increased discrepancies with

observations. The large uncertainty in alb, both in winter and summer (Fig. 8), is related to a possible overestimation of the nutrient inflow through Gibraltar Strait. In general, Fig.8 shows that BIAS is positive in winter for the western sub-basins, while is almost negligible in summer for all sub-basins. The value of the chlorophyll RMSD over the Mediterranean Sea, considering the 2016-2017 average, is 0.045 and 0.015 mg/m^3 for winter and summer, respectively, while BIAS is 0.015 and

5 -0.005 mg/m^3 in winter and in summer. The recently upgraded assimilation scheme that integrates both coastal and open-sea chlorophyll data (Teruzzi et al., 2018a) provides a good model performance also in the coastal areas. In these areas the model underestimates the satellite product of about 0.1 mg/m^3 in both seasons, and the mean RMSD is about 0.4 mg/m^3, with higher values (between 0.5 and 0.9 mg/m^3) in areas strongly influenced by coastal processes (not shown). Uncertainty in model prediction in coastal areas is mostly related to the lack of high frequency data for river nutrient discharges which limits the

10 model capability to simulate bloom events triggered by rivers plume events (Teruzzi et al., 2018a).

The comparison of model chlorophyll output with the BGC-Argo floats (Fig. 9 and Tab. 3) provides a skill performance analysis of the model quality in reconstructing the vertical dynamics, integrating the assessment on model surface performances. The Hovmoller diagrams of Fig. 9 show how the time evolution of the model vertical profiles matches up the observations along the corresponding float trajectory. The very good qualitative agreement of the MedBFM model with the

15 BGC-Argo floats is highlighted by the consistent temporal succession of the winter vertically mixed blooms, the onset, the time evolution and the depth of the deep chlorophyll maximum (DCM), which typically establishes during the stratified season. The time series of the quantitative metrics (defined in Sect. 4.1) computed on the vertical profiles comparison are shown in the right column plots of Fig. 9 for the selected model-float pairs. The agreement between model and float chlorophyll at the surface and its vertical average in the 0-200 m layer is fairly good with a slight underestimation of the 0-200 m integrated

20 values during winter. Correlation values of the selected float are almost always larger than 0.7, higher during summer and lower in winter. The DCM depth is very well captured by the MedBFM, both in terms of vertical displacement and temporal evolution for the two selected floats. The depth of MWB shows some inconsistency between model and float data, but this metric is not completely effective since it is not always computable from BGC-Argo floats data.

Averaging the time series of RMSD and BIAS of the new metrics for the aggregated sub-basins (Tab. 3) highlights that the

25 MedBFM model has a very high skill in reproducing the vertical dynamics of the phytoplankton chlorophyll in the 0-200 m layer, both considering the very high spatial heterogeneity of the Mediterranean Sea and the seasonal cycle of the coupled physical-biogeochemical processes. In particular, the correlation between vertical profiles of model and observation ranges from 0.7 to 0.85, with the exception of the Alboran Sea (where only 2 profiles per month are available). The uncertainty of the DCM position is less than 20 m with a BIAS between -9 and 7 m for the areas with at least 10 float profiles per month, which

30 is a very inspiring result considering that the model vertical discretization is about 6-8 m for the layers around the depth of 80-120 m.

The averaged vertical values show that the model generally underestimates the content of chlorophyll with respect the BGC-Argo floats measurements, which however appears in contrast with the general assessment of model overestimation for the winter period w.r.t. the satellite data. Triple collocation method, as proposed by Mignot et al. (2018), might be applied to

investigate possible off-sets and random errors among multi-platform datasets at regional/local scale. Nevertheless, the RMSD of the 0-200 m vertical averages remains lower than 0.1 mg/m^3 in all the aggregated sub-basins. The depth of the vertically mixed winter bloom (MWB) is not computable and reliable for some of the sub-basins, however for those having at least 10 float profiles per month, the MWB has an absolute bias ranging from 30 to 40 m and a RMSD ranging from 40 to 50 m.

5 The comparison of model nitrate with the BGC-Argo float measurement allows to evaluate the skill of the MedBFM to simulate key coupled physical-biogeochemical processes (i.e., water column nutrient content, nitracline and effect of winter mixing and summer stratification on the shape of nitrate profile; metrics defined in Sect. 4.1). Qualitatively, we observe a general good model performance in simulating the shape of the profile (i.e. correlation values), the temporal evolution of the 0-200 m averaged values and of the nitracline depth of the selected float (Fig. 10). The nitrate metrics of the 8 floats are then averaged

10 over the aggregated sub-basins (Tab. 4): even if the scarcity of the profiles possibly limits the generalization of the results, our validation framework highlights that the MedBFM model system shows excellent performance in simulating the shape of profiles and the seasonal evolution of the mesoscale dynamics affecting the nitrate field. In particular, Tab. 4 reports that the mean value of nitrate on the 0-200 m layer is very well simulated, with BIAS ranging from 0.04 to -0.68 mmol/m^3 and RMSD generally smaller than 1 mmol/m^3; the correlation is always higher than 0.9 and the depth of the nitracline is simulated with

15 an uncertainty lower than 40 m. Further, accordingly with BGC-Argo floats observations (Tab. 4), the MedBFM reproduces fairly well the Mediterranean basin scale heterogeneity with a nitracline at around 60-100 m in the western sub-basins and below 110 m in the eastern sub-basins (with absolute BIAS never larger than 35 m and uncertainty between 20 and 40 m).

The qualitative comparison of modelled oxygen with a selected BGC-Argo float (Fig. 11) shows the MedBFM skill to simulate the sequence of physical-biogeochemical processes of the oxygen dynamics, such as the effect of ventilation during winter,

20 the production of an oxygen maximum at the layer of the DCM due to the intense phytoplankton production during spring and summer, and the minimum of oxygen concentration at surface during summer and autumn due to decrease of saturation and presence of consumption terms. Interestingly, the depth of ventilation has a clear interannual variability, as shown by the higher values of oxygen below the 100 m depth in the event of December 2017 – January 2018 with respect to the previous year. The quantitative comparison between all the available floats data and model results is summarized by the statistics of Tab. 5,

25 showing a general model overestimation of about 15 mmol/m^3 at surface, increasing with depth to about 20-25 mmol/m^3. The Adriatic Sea (data from 1 float only) shows a much lower discrepancy of around 5-10 mmol/m^3.

Discrepancies at surface might be due to saturation calculation, whereas at depth to inaccuracies of the initial conditions or to excess of production. However, considering the modelled bias error in temperature and salinity at surface of -0.23° and 0.01, respectively (Clementi et al., 2018), and under the hypothesis of oxygen saturation at surface, the BIAS for the modelled

30 oxygen (i.e., calculated using the formulations of Weiss, 1970, and of Garcia and Gordon, 1992) should not exceed 1-1.5 mmol/m^3 throughout the year. On the other hand, the on-going improvement of quality control procedures (Johnson et al., 2017) and the need for reprocessing might have an impact on the accuracy of archived oxygen data. Thus, this comparison must be considered cautionary, nevertheless it provides a qualitative indication of the model behaviour to capture spatial and temporal oxygen dynamics.

5.2 Near-real time forecast skill performance

The near-real time (NRT) skill performance of the operational forecast system aims at delivering sustained on-line information on the quality of Med-BIO biogeochemical forecast products, i.e., firstly identifying main biases and possible suspicious trends in the forecasts, and secondly establishing that the accuracy remains within the assessed ranges. The NRT validation activities

5   are performed using GODAE Class 4 metrics with available satellite data for the first three days of forecast (T0-T2 lead time) and BGC-Argo floats observations for the first four days of forecast (T0-T3 lead time) using the same metrics computed for the pre-operational run, that provides the benchmarks of the accuracy level. Online resources for this metrics are updated quarterly on the official CMEMS validation webpage[5] and weekly on the regional Mediterranean validation website managed by OGS[6].

10   Figure 12 reports the RMSD between NRT daily L3 multi-sensor satellite data (see details in Sect. 2.2) and the first three days of forecast for selected sub-basins since April 2018 (i.e., the start of the last version of the CMEMS Med-BIO system at the time of writing). Similarly to the pre-qualification run (Fig. 8), the forecast skill metrics are characterized by a seasonal and spatial variability that basically reflects the chlorophyll spatial and temporal variability. For the period reported in Fig. 12, the performance of the first day of forecast is generally better than the benchmark references, while it decreases for the second

15   and third day of forecast. Indeed, the average of the RMSD over the 16 sub-basins is 0.018, 0.034 and 0.41 mg/m^3 for the first, second and third day of forecast, respectively. The high variability of the RMS statistics from one day to another is basically due to the daily varying number of available pixels due to the cloud cover.

Figure 13 shows the distribution of the available BGC-Argo data matched up with the forecast data of chlorophyll, nitrate and oxygen basin-averaged on different vertical layers for the first four days of forecast, and a season-based benchmark represented

20   by the results from the pre-operational run.

In general, the forecast data are generally within the variability of the seasonal benchmark (May to August, orange points). Indeed, the overall RMSD metrics of the forecast skill of chlorophyll and nitrate are always lower than the values estimated for the pre-operational run (Tab. 6), while the RMSD statistics of oxygen forecast highlight the bias in the lower layers and are slightly higher than the computed RMSD for the pre-operational run. No significant differences can be recognized from

25   the distribution of the four days of forecast points. Correspondingly, a trend of the skill performance among the four days of forecast is not clearly visible, with the exception of nitrate RMSD that decreases from T0 to T4, while chlorophyll increases from T1 to T4. In any case, the very low number of available data (few tens in the 5 months considered) does not allow to compute significant statistics on the persistency of the quality among the forecast days.

---

[5] Available at http://marine.copernicus.eu/services-portfolio/scientific-quality/
[6] Available at http://medeaf.inogs.it/nrt-validation/

6 Discussion

This work presents the last achievements in the operational biogeochemical component for the Mediterranean Sea delivered by CMEMS. The MedBFM model system has been integrated with the last scientific achievements of the BFM model (Cossarini et al., 2015; Lazzari et al., 2016), the 3DVarBio assimilation scheme (Teruzzi et al., 2018a,b) and the non-linear
5 free surface and volume vertical layer parameterization of the transport operator of the OGSTM model (Salon et al., 2018).
The Med-BIO system has followed the developments of the EU operational marine services (Le Traon et al., 2017), starting from its first version (Lazzari et al., 2010) deployed within the MERSEA project (2004-2008; GMES implementation phase), becoming pre-operational during MyOcean projects series (2009-2015; GMES demonstration and pre-operational phase) and finally establishing a regular and validated operational product delivery in CMEMS (GMES operational phase). Across this
10 10-year period, the quality of the Med-BIO products has significantly increased (Fig. 14, quality assessed by the RMSD of the surface chlorophyll concentration, the only product variable that has been consistently validated since the beginning of the Med-BIO activity), with a continuous improvement which took advantage from the implementation of the data assimilation, from the increased horizontal resolution, from the evolutions in the physical component of the Med-MFC system.
The Med-BIO off-line coupling with Med-PHY was outlined since the preliminary work of Lazzari et al. (2010) and has
15 allowed for distinctive developments of the different components. Further, the alignment between physical and biogeochemical models in terms of same horizontal resolution, bathymetry, boundaries (number and position of rivers) and surface forcing (e.g., z* parameterization), a requisite of the CMEMS framework, guarantees the consistency of the results (as shown by the recent improvement of the performance after April 2018, Fig. 14). Other studies demonstrated that off-line coupling does not affect the transport of biogeochemical tracers when the sub-mesoscale physics is degraded to mesoscale (Levy et al., 2012).
20 Further improvements of the Med-BIO biogeochemical model system in term of physical-biogeochemical consistency at local scale are expected with the foreseen implementation of the assimilation of the BGC-Argo floats data (Cossarini et al., 2019), which has shown the improvement of the model solution due to the increased consistency of vertical dynamics by the assimilation of the physical and biogeochemical profiles at the same time and position. This result highlights the importance of the joint physical and biogeochemical assimilation, that has been recently demonstrated in a twin experiment to provide
25 superior results with respect to any uncoupled assimilation configuration (Yu et al., 2018).

Communicating the uncertainty is a critical point: it helps the users to properly interpret the validity of the forecast products, even when the forecast actually fails, and to minimize any problem created by the misuse (and misinterpretation) of them (Stow et al., 2009; Payne et al., 2017). The communication of the level of uncertainty of the biogeochemical product remains
30 an open issue for the scarcity of reference NRT biogeochemical observations available and for the complexity of biogeochemical models, which may have tens of variables but only a few can be validated. Further, regional operational models have reached the limit of the sub-mesoscale, which is not adequately sampled by observational systems (Hernandez et al., 2018). As an example, the number of dissolved oxygen observations used to build Fig. 6 is almost one fifth of those available

for phosphate (Cossarini et al., 2018), therefore the reliability of validation using the climatological profiles might be lower and even less for the surface values, since dissolved oxygen exhibits a significant seasonal and high frequency cycles due to the air-sea exchanges mediated by saturation.

We show that depending on the variables, different uncertainty levels can be provided on the basis of the availability of reference data. In this context, the validation analysis provides a "degree of confirmation" (Oreskes et al., 1994) with respect to the different scales of variability derived from the available observations. GODAE Class 1 metrics show that the model is consistent (in terms of chlorophyll, nitrate, phosphate, oxygen, dissolved inorganic carbon, alkalinity and pH) in reproducing the vertical profile climatology, at sub-basin scale and for the sixteen sub-basins (Figs. 3 to 8). The comparison of model primary production with available basin-wide estimates and literature collection can only provide a consistency confirmation of the model estimates at the basin and annual scales. Then, we also demonstrate that GODAE Class 4 metrics are feasible and provide more rigorous skill performance down to the scale of week and mesoscale, but only for a limited number of variables (see Figs. 9 to 13). Regarding data availability, satellite chlorophyll estimates (Colella et al., 2016) represent the most reliable source of NRT data, which, however, allows to investigate only the cloud-free surface of the ocean. The novel BGC-Argo floats data set empowered us to design new skill metrics, showing the capability of the MedBFM model to reproduce the temporal evolution of the vertical dynamics of the phytoplankton, nitrate and oxygen, and to assess key ecosystem processes in the Mediterranean Sea. The novel metrics based on BGC-Argo data disclose new and important perspectives for the model validation in the Mediterranean Sea, also considering its very high spatial heterogeneity and the seasonal variability of the coupled physical-biogeochemical processes. However, some cautions should be taken before generalizing the conclusions, since the relatively poor BGC-Argo floats coverage in some areas and the on-going improvement of product quality procedures of the BGC-Argo data (Johnson et al., 2017). Moreover, the relevance of the representation error (Hernandez et al., 2018) becomes stricter with BGC-Argo data, since the skill performance analysis is based on comparing a model output with grid cells of 3-4 km^2 wide and O(10) meter thick with point-like profiles of few meters of resolution, thus the model may miss part of the spatial-temporal scales present in the observations (Oke and Sakov, 2008).

Considering the NRT evaluation, Figures 12 and 13 show that the uncertainty of the forecast products is of the same order (and in some occasions even lower) than the pre-operational run, and that the performance decreases slightly from the first day of forecast to the following ones. This is due to both the uncertainty due to the intrinsic error of the biogeochemical model and the decrease of the performance of the physical variable forecasts driving the biogeochemical ones (see as reference, Clementi at al., 2018, and the CMEMS quarterly validation statistics[7]). Furthermore, comparing NRT metrics with the benchmark can highlight anomalous model behaviors that may constitute an operational monitoring system to alert the users and the researcher staff that model performance is worsening or a specific event is occurring, thus conveying useful information to investigate possible causes.

---

[7] Available at http://marine.copernicus.eu/services-portfolio/scientific-quality/.

We show that the error statistics, in terms of RMSD, are proportional to the variability of the variables. It is shown for surface chlorophyll (Fig. 8) and it can be also noted from BGC-Argo derived statistics: sub-basins characterized by higher variability have higher error (Tab. 3). As a result, the performance analysis shows that the western regions have, in general, largest variability and lower performance, specifically during the winter season. Thus, to rationalize the costs of observing systems

5   (Cristini et al., 2016), it may be more efficient to sustain the observing systems with high-frequency observations in high-variability areas. On the other hand, given that fields variability may be related to local physical and biogeochemical processes (e.g. vertical mixing, coastal effects due to strong topographic gradients or terrestrial inputs), the reduction of the model representativeness error can benefit from a more collaborative evolution of the coupled physical-biogeochemical systems, both in terms of process modeling or coupled data assimilation.

10   The present validation framework uses an *a priori* subdivision which considers the biogeographic approach of D'Ortenzio and Ribera d'Alcalà (2009), and subsequent refinement proposed by Lazzari et al. (2012), which showed different characterizations according to Longhurst paradigm. The recent review of Ayata et al. (2018) discusses the variations of the Mediterranean Sea subdivision found in literature, highlighting regions with relatively homogeneous conditions and some heterogeneous regions featuring significant mesoscale activity. Our validation approach demonstrates the importance to provide model uncertainty

15   estimation at different spatial and temporal scales, emphasizing the model capability to reproduce specific processes and their intensity in different areas, while computing metrics over the sub-regions allows to synthetize the heterogeneity of the Mediterranean Sea. The use of simple indexes, such as means and standard deviation, and functional spatio-temporal subdivisions increase the readability of the uncertainty communication, which responds also to the request for a user oriented evolution of the valuation framework in operational systems (Hernandez et al., 2018).

20   Our validation results point out a number of strengths and weaknesses of the CMEMS Mediterranean forecasting biogeochemical system. A strength is that the MedBFM is operationally in place and provides validated and reliable ecosystem products consistently with the physical ones (Clementi et al., 2017).

The system can also provide important feedbacks to the observing autonomous systems. Indeed, the NRT comparison of BGC-Argo floats data with the forecast outputs w.r.t to the benchmarking might be beneficial for an additional QC procedure for

25   detecting anomalous observations that the present QC fails to detect, as proposed for physical variables measured by Argo systems (Ingleby and Huddleston, 2007). As an example, the Class 4 metrics applied to BGC-Argo oxygen (Fig. 13, third column), shows a systematic bias which does not appear when contrasted with the Class 1 validation (Figs. 5 and 7), thus pointing out the opportunity of a possible revision of some model formulations or product quality procedure of BGC-Argo oxygen in the Mediterranean Sea.

30   Another positive aspect of our work is that, to the best of our knowledge, this is one of the first times that a consistent validation procedure provides sustainable guidelines following GODAE metrics for operational marine biogeochemistry exploiting the BGC-Argo floats data (Hernandez et al., 2018). Our novel metrics (Figs. 9 and 10) provided indications of the model skill performance on some key biogeochemical processes (DCM, nutricline depth), thus setting an advancement to what described

in Hernandez et al. (2015) for NRT assessment of biogeochemical operational forecast and maximizing the values of the available NRT biogeochemical observations (She et al., 2016).

On the other hand, a weakness of MedBFM is the reduction in performance close to the domain boundaries at Gibraltar and Dardanelles Straits and in the coastal areas. The observed overestimation of chlorophyll, and thus of productivity and phytoplankton biomass, in Alboran Sea (see Figs. 3 and 8) can be related to an incorrect parameterization of the biogeochemical fluxes through Gibraltar Strait or to the effect of vertical mixing. Inconsistent physical-biogeochemical data assimilation might generate incompatible density and nutrients profiles that may generate an extra amount of vertical flux of nutrient in this highly dynamical area, thus enhancing its productivity. Increase of nutrient availability along isocline surfaces has been observed by Raghukuma et al. (2015) suggesting this as a possible cause of increase of productivity in oligotrophic areas. The upgrade of MedBFM boundary conditions (at the Gibraltar Strait, Med-PHY is coupled with the CMEMS global product while Med-BIO uses climatological biogeochemical value) with high frequency values, and the extension of the Atlantic buffer zone, could improve the model performance in this area.

For the coastal areas, the increased resolution to 1/24° cannot fully balance the use of low frequency data of biogeochemical terrestrial inputs (i.e. nutrients and carbonate system estimates from climatological databases). Thus, some quality decrease is observed even though data assimilation of coastal chlorophyll from satellite can partly reduce this deficiency (Teruzzi et al., 2018a). Operational or at least higher frequency coastal data for rivers and the inclusion of Dardanelles as an open boundary condition are requested to account for the user needs of reliable products in coastal areas.

7 Conclusions

The present work evaluates the skill performance of the CMEMS Mediterranean Biogeochemistry component (Med-BIO) determining the quality of the CMEMS biogeochemical products on the basis of two complementary phases: 1) the pre-operational qualification run (2016-2017), and 2) the operational workflow (started in April 2018 for MedBFMv2.1).

Using different observation reference data sets (from satellite, literature, climatology, BGC-Argo floats), GODAE Class 1 and 4 metrics have been applied to the MedBFM model system in order to quantify its consistency in simulating the key features of the Mediterranean biogeochemistry, and its accuracy to routinely reproduce the observations at their specific time and locations. New metrics specifically designed to exploit the richness of BGC-Argo floats database and to evaluate the model capability to reproduce the key elements of the vertical profiles of chlorophyll and nitrate have been proposed. Main results can be here summarized:

- MedBFM is consistent in reproducing the general characteristics of biogeochemistry in Mediterranean Sea, and the CMEMS Med-BIO products are well within the climatological variability; quantified correlation values are larger than 0.9 and 0.7 for nutrients and carbonate system products, respectively.
- The level of accuracy of the different Med-BIO products depends on the kind of variable, the availability of reference data, the sub-basin and the season.

- Novel Class 4 metrics based on the model match-up with BGC-Argo floats data represent an useful tool to quantify the capability of a biogeochemical model to reproduce key elements of the biogeochemical processes along the water column (depth of deep chlorophyll maximum, mixed winter bloom, nutricline). For MedBFM, correlation is generally larger than 0.7/0.9 for vertically integrated chlorophyll/nitrate, and errors (as RMSD) in reproducing the key depths ranges between 12 and 50 m.

- NRT validation of Med-BIO forecast products have been performed for chlorophyll, nitrate and oxygen from April 2018, showing a decrease of forecast skill performance after 1 or 2 days for surface chlorophyll and a not clearly identified pattern when BGC-Argo data are used. Nevertheless, the forecast skill performance remains at the same level as the benchmark.

Even if the use of BGC-Argo floats significantly discloses new perspectives for operational biogeochemical model validation, some cautions should be considered before generalizing the conclusions, due to the relatively poor BGC-Argo coverage in some areas of the Mediterranean Sea and the on-going improvement of product quality procedures of the BGC-Argo data. Finally, the validation metrics here presented provides indications of some weaknesses of the Med-BIO (e.g. limited dynamics in coastal areas, Gibraltar boundary and sub-mesoscale effects on phytoplankton dynamics in western area) that will lead to future developments. Nevertheless, the validation results support not only the accuracy of the CMEMS Med-BIO products, but also the consistency of the MedBFM model system to simulate the fundamental coupled physical-biogeochemical processes, which is corroborated at the mesoscale and weekly scale.

Author contribution

SS and GC conceived the ideas of the work, the formulation of main research goals, the investigation, and the methodology. SS was in charge of the supervision of the manuscript preparation, preparing the original draft with contributions from all co-authors. GC and SS conducted the formal analysis supported by LF. GB was in charge of the data curation, LF prepared most of the figures. GC, PL, AT and LF contributed to review the original draft.

GC is responsible of the CMEMS Med-BIO system; GB is in charge of the operational workflow; LF conducts the NRT operational validation; GC, GB, AT, SS and PL worked on the upgrade of the CMEMS Med-BIO system. CS, AC, GC and SS contributed to the funding acquisition.

Acknowledgements

This study has been conducted using E.U. Copernicus Marine Service Information.

References

Ayata, S. D., Irisson, J. O., Aubert, A., Berline, L., Dutay, J. C., Mayot, N., Nieblas, A.-E., D'Ortenzio, F., Palmièri, J., Reygondeau, G., and Rossi, V.: Regionalisation of the Mediterranean basin, a MERMEX synthesis, Prog. Oceanogr., 163, 7-20, https://doi.org/10.1016/j.pocean.2017.09.016, 2018

5 Bell, M.J., Schiller, A., Le Traon, P.-Y., Smith, N.R., Dombrowsky, E., and Wilmer-Becker, K.: An introduction to GODAE OceanView, J. Oper. Oceanogr., 8, s2-s11, http://dx.doi.org/10.1080/1755876X.2015.1022041, 2015.

Berx B., Dickey-Collas M., Skogen M.D., de Roeck Y.H., Klein H., Barciela R., Forster R.M., Dombrowsky E., Huret M., Payne M., Sagarminaga, Y., And Schrum, C.: Does Operational Oceanography Address the Needs of Fisheries and Applied Environmental Scientists?, Oceanography, 24, 166–171, doi: 10.5670/oceanog.2011.14, 2011.

10 Bolzon, G., Cossarini, G., Lazzari, P., Salon, S., Teruzzi, A., Crise, A., and Solidoro, C.: Mediterranean Sea biogeochemical analysis and forecast (CMEMS MED AF-Biogeochemistry 2013-2017), Copernicus Monitoring Environment Marine Service, https://doi.org/10.25423/MEDSEA_ANALYSIS_FORECAST_BIO_006_006, 2017.

Clementi, E., Oddo, P., Drudi, M., Pinardi, N., Korres, G., and Grandi, A.: Coupling hydrodynamic and wave models: first step and sensitivity experiments in the Mediterranean Sea, Ocean Dynam, 67, 1293-1312, https://doi.org/10.1007/s10236-017-

15 1087-7, 2017.

Clementi, E., Grandi, A., Di Pietro, P., Pistoia, J., and Mattia, G.: Quality Information Document for MEDSEA_ANALYSIS_FORECAST_PHY_006_013, Copernicus Marine Environment Monitoring Service, available at: http://cmems-resources.cls.fr/documents/QUID/CMEMS-MED-QUID-006-013.pdf, 2018.

Colella, S.: La produzione primaria nel Mar Mediterraneo da satellite: sviluppo di un modello regionale e sua applicazione ai

20 dati SeaWiFS, MODIS e MERIS, Phd Thesis, Università Federico II, Napoli, 162 pp., 2006.

Colella, S., Falcini, F., Rinaldi, E., Sammartino, M., and Santoleri, R.: Mediterranean Ocean Colour Chlorophyll Trends. PLoS ONE, 11, e0155756. https://doi.org/10.1371/journal.pone.0155756, 2016.

Cossarini, G., Lazzari, P., and Solidoro, C.: Spatiotemporal variability of alkalinity in the Mediterranean Sea. Biogeosciences, 12, 1647-1658, https://doi.org/10.5194/bg-12-1647-2015, 2015.

25 Cossarini, G., Salon, S., Bolzon, G., Teruzzi, A., Lazzari, P., and Feudale, L.: Quality Information Document for MEDSEA_ANALYSIS_FORECAST_BIO_006_014, Copernicus Monitoring Environment Marine Service, available at: http://cmems-resources.cls.fr/documents/QUID/CMEMS-MED-QUID-006-014.pdf, 2018.

Cossarini, G., Mariotti, L., Feudale, L., Teruzzi, A., D'Ortenzio, F., Tallandier, V., and Mignot, A.: Towards operational 3D-Var assimilation of chlorophyll Biogeochemical-Argo float data into a biogeochemical model of the Mediterranean Sea, Ocean

30 Model., 133, 112-128, https://doi.org/10.1016/j.ocemod.2018.11.005, 2019.

Cristini, L., Lampitt, R. S., Cardin, V., Delory, E., Haugan, P., O'Neill, N., Petihakis, G., and Ruhl, H. A.: Cost and value of multidisciplinary fixed-point ocean observatories, Marine Policy, 71, 138-146, https://doi.org/10.1016/j.marpol.2016.05.029, 2016.

Dobricic, S., and Pinardi, N.: An oceanographic three-dimensional variational data assimilation scheme, Ocean Model, 22, 89-105, https://doi.org/10.1016/j.ocemod.2008.01.004, 2008.

D'Ortenzio, F., and Ribera d'Alcalà, M.: On the trophic regimes of the Mediterranean Sea: a satellite analysis, Biogeosciences, 6, 139-148, https://doi.org/10.5194/bg-6-139-2009, 2009.

5   Foujols, M.-A., Lévy, M., Aumont, O., and Madec, G.: OPA 8.1 Tracer Model Reference Manual, Institut Pierre Simon Laplace, France, 39 pp., 2000.

Gehlen, M., Barciela, R., Bertino, L., Brasseur, P., Butenschön, M., Chai, F., Crise, A., Drillet, Y., Ford, D., Lavoie, D., Lehodey, P., Perruche, C., Samuelsen, A., and Simon, E.: Building the capacity for forecasting marine biogeochemistry and ecosystems: recent advances and future developments, J. Oper. Oceanogr., 8, s168-s187,

10   https://doi.org/10.1080/1755876X.2015.1022350, 2015.

Garcia, H. E., Gordon, L. I.: Oxygen solubility in seawater: Better fitting equations. Limnology and oceanography, 37(6), 1307-1312, 1992.

Gregg, W. W., Friedrichs, M. A., Robinson, A. R., Rose, K. A., Schlitzer, R., Thompson, K. R., and Doney, S. C.: Skill assessment in ocean biological data assimilation, J. Marine Sys., 76, 16-33, https://doi.org/10.1016/j.jmarsys.2008.05.006,

15   2009.

Günther, H., and Behrens, A.: The WAM model. Validation document Version 4.5.4, Institute of Coastal Research Helmholtz-Zentrum Geesthacht, Germany, 2012.

Hernandez, F., Blockley, E., Brassington, G.B, Davidson, F., Divakaran, P., Drévillon, M., Ishizaki, S., Garcia-Sotillo, M., and other 17 co-authors: Recent progress in performance evaluations and near real-time assessment of operational ocean

20   products, J. Oper. Oceanogr., 8, 221-238, https://doi.org/10.1080/1755876X.2015.1050282, 2015.

Hernandez, F., Smith G., Baetens K., Cossarini G., Garcia-Hermosa I., Drévillon M., Maksymczuk J., Melet A., Régnier C., and von Schuckmann K.: Measuring performances, skill and accuracy in operational oceanography: New challenges and approaches, in: New Frontiers in Operational Oceanography, edited by: Chassignet, E., Pascual, A., Tintoré, J., and Verron, J., GODAE OceanView, 759-796, https://doi.org/10.17125/gov2018, 2018.

25   Ingleby, B., and Huddleston, M.: Quality control of ocean temperature and salinity profiles—Historical and real-time data, J. Marine Sys., 65, 158-175, https://doi.org/10.1016/j.jmarsys.2005.11.019, 2007.

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

30   Johnson, K.S., and Coletti, L.J: In situ ultraviolet spectrophotometry for high resolution and long-term monitoring of nitrate, bromide and bisulfide in the ocean, Deep-Sea Res. Pt. I, 49, 1291-1305, https://doi.org/10.1016/S0967-0637(02)00020-1, 2002.

Lazzari, P., Teruzzi, A., Salon, S., Campagna, S., Calonaci, C., Colella, S., Tonani, M., and Crise A.: Pre-operational short-term forecasts for the Mediterranean Sea biogeochemistry, Ocean Sci., 6, 25-39, https://doi.org/10.5194/os-6-25-2010, 2010.

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

Lazzari, P., Solidoro, C., Salon, S., and Bolzon, G.: Spatial variability of phosphate and nitrate in the Mediterranean Sea: a modelling approach, Deep-Sea Res. Pt. I, 108, 39-52, https://doi.org/10.1016/j.dsr.2015.12.006, 2016.

Lavigne, H., D'ortenzio, F., d'Alcalà, M.R,. Claustre, H., Sauzède R., and Gacic, M.: On the vertical distribution of the chlorophyll a concentration in the Mediterranean Sea: a basin-scale and seasonal approach, Biogeosciences, 12, 5021-5039, https://doi.org/10.5194/bg-12-5021-2015, 2015.

Le Traon, P.Y., and other 100 co-authors: The Copernicus Marine Environmental Monitoring Service: Main Scientific Achievements and Future Prospects, Special Issue Mercator Océan Journal, 56, https://doi.org/10.25575/56, 2017.

Lewis, E., and Wallace, D.W.R.: Program Developed for CO2 System Calculations, ORNL/CDIAC-105, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, available at: http://cdiac.ornl.gov/oceans/co2rprt.html, 1998.

Macias, D. , Garcia-Gorriz, E. and Stips, A.: Major fertilization sources and mechanisms for Mediterranean Sea coastal ecosystems. Limnol. Oceanogr., 63: 897-914. doi:10.1002/lno.10677, 2018.

Madec, G. and the NEMO team: NEMO ocean engine: v3.6 stable, Note du Pôle de modélisation de l'Institut Pierre-Simon Laplace, France, 27, 1288-1619, 2016.

Mattern, J.P., Edwards, C.A., and Moore, A.M.: Improving variational data assimilation through background and observation error adjustments, Mon. Weather Rev., 146, 485-501, https://doi.org/10.1175/MWR-D-17-0263.1, 2018.

Melaku Canu, D., Ghermandi, A., Nunes, P.A.L.D., Cossarini, G., Lazzari, P., and Solidoro, C.: Estimating the value of carbon sequestration ecosystem services in the Mediterranean Sea: An ecological economics approach, Global. Environ. Chang., 32, 87-95, https://doi.org/10.1016/j.gloenvcha.2015.02.008, 2015.

Mignot, A., D'Ortenzio, F., Taillandier, V., Cossarini, G., and Salon, S.: Quantifying observational errors in Biogeochemical-Argo oxygen, nitrate and chlorophyll a concentrations, Geophys. Res. Lett., in review. 2018.

Oke P.R., and Sakov P.: Representation error of oceanic observations for data assimilation, J. Atmos. Ocean. Tech., 25, 1004–1017, https://doi.org/10.1175/2007JTECHO558.1, 2008.

Oreskes, N., Shrader-Frechette, K., and Belitz, K.: Verification, validation, and confirmation of numerical models in the earth sciences, Science, 263, 641-645, http://doi.org/10.1126/science.263.5147.641, 1994.

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 Sci., 5, 461-473, https://doi.org/10.5194/os-5-461-2009, 2009.

Oddo, P., Bonaduce, A., Pinardi, N., and Guarnieri, A.: Sensitivity of the Mediterranean sea level to atmospheric pressure and free surface elevation numerical formulation in NEMO. Geosci. Model Dev., 7, 3001–3015, https://doi.org/10.5194/gmd-7-3001-2014, 2014.

Payne, M.R., Hobday, A.J., MacKenzie, B.R., Tommasi, D., Dempsey, D.P., Fässler, S.M.M., Haynie, A.C., Ji, R., Liu, G., Lynch, P.D., Matei, D., Miesner, A.K., Mills, K.E., Strand, K.O., and Villarino, E.: Lessons from the First Generation of Marine Ecological Forecast Products, Front. Mar. Sci., 4, 289, https://doi.org/10.3389/fmars.2017.00289, 2017.

Pasqueron de Fommervault, O., D'Ortenzio, F., Mangin, A., Serra, R., Migon, C., Claustre, H., Lavigne, H., d'Alcalà, M.R.,
5   Prieur, L., Taillandier, V., and Schmechtig, C.: Seasonal variability of nutrient concentrations in the Mediterranean Sea: Contribution of Bio-Argo floats, J. Geophys. Res.-Oceans, 120, 8528-8550, https://doi.org/10.1002/2015JC011103, 2015.

Pistoia, J., Clementi, E., Delrosso, D., Mattia, G., Fratianni, C., Drudi, M., Grandi, A., Padeletti, D., Di Pietro, P., Storto, A., and Pinardi, N.: Last improvements in the data assimilation scheme for the Mediterranean Analysis and Forecast system of the Copernicus Marine Service, in: Operational Oceanography serving Sustainable Marine Development, Proceedings of the
10 Eight EuroGOOS International Conference, Bergen, Norway, 3-5 October 2017, 335-342, 2018.

Roesler, C., Uitz, J., Claustre, H., Boss, E., Xing, X., Organelli, E., Briggs, N., Bricaud, A., Schmechtig, C., Poteau, A., D'Ortenzio, F., Ras, J., Drapeau, S., Haëntjens, N., and Barbieux, M.: Recommendations for obtaining unbiased chlorophyll estimates from in situ chlorophyll fluorometers: A global analysis of WET Labs ECO sensors, Limnol. Oceanogr-Meth., 15, 572–585, https://doi.org/10.1002/lom3.10185, 2017.

15 Ryan, A.G., Regnier, C., Divakaran, P., T. Spindler, Mehra, A., Smith, A.G., Davidson, F., Hernandez, F., Maksymczuk, J. and Liu, Y.: GODAE OceanView Class 4 forecast verification framework: global ocean inter-comparison, J. Oper. Oceanogr., 8, s98-s111, https://doi.org/10.1080/1755876X.2015.1022330, 2015.

Salon, S., Cossarini, G., Lazzari, P., Teruzzi, A., Di Cerbo, P., Bolzon, G., Feudale, L., Solidoro, C., and Crise, A.: High resolution operational analysis and forecasts for the Mediterranean Sea biogeochemistry. Operational Oceanography serving
20 Sustainable Marine Development, in: Proceedings of the Eight EuroGOOS International Conference, Bergen, Norway, 3-5 October 2017, 285-294, 2018.

Schmechtig, C., and Thierry, V.,; Argo quality control manual for biogeochemical data, The Bio Argo Team, available at: http://doi.org/10.13155/40879, 2016.

She, J., Allen, I., Buch, E., Crise, A., Johannessen, J. A., Le Traon, P.-Y., Lips, U., Nolan, G., and other 5 co-authors:
25 Developing European operational oceanography for Blue Growth, climate change adaptation and mitigation, and ecosystem-based management, Ocean Sci., 12, 953-976, https://doi.org/10.5194/os-12-953-2016, 2016.

Storto, A., Masina, S., and Navarra, A.: Evaluation of the CMCC eddy-permitting global ocean physical reanalysis system (C-GLORS, 1982-2012) and its assimilation components, Q. J. Roy. Meteor. Soc., 142, 738–758, https://doi.org/10.1002/qj.2673, 2015.

30 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: Qc Procedure for Profiling Float Oxygen, J. Geophys. Res.-Oceans, 118, 5640–5650. https://doi.org/10.1002/jgrc.20399, 2013.

Teruzzi, A., Salon, S., Bolzon, G., Lazzari, P., Ficarelli, F., Solidoro, C., and Crise, A.: Operational forecasts of the biogeochemical state of Mediterranean Sea, Mercator Ocean Quarterly Newsletter, 40, 15-25, available at: https://www.mercator-ocean.fr/wp-content/uploads/2015/05/Mercator-Ocean-newsletter-2011_40.pdf, 2011.

Teruzzi, A., Dobricic, S., Solidoro, C., and Cossarini, G.: A 3D variational assimilation scheme in coupled transport biogeochemical models: Forecast of Mediterranean biogeochemical properties, J. Geophys. Res.-Oceans, https://doi.org/10.1002/2013JC009277, 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 modelling in the Mediterranean Sea, Ocean Model., 132, 46-60, https://doi.org/10.1016/j.ocemod.2018.09.007, 2018a.

Teruzzi, A., Di Cerbo, P., Cossarini, G., Pascolo, E., and Salon, S.: Parallel implementation of a data assimilation scheme for operational oceanography: the case of the OGSTM-BFM model system, Computers and Geosciences, in review, 2018b.

Tonani, M., Pinardi, N., Dobricic, S., Pujol, I., and Fratianni, C.: A high-resolution free-surface model of the Mediterranean Sea, Ocean Sci., 4, 1-14, https://doi.org/10.5194/os-4-1-2008, 2008.

Tonani, M., Nilsson, J. A. U., Lyubartsev, V., Grandi, A., Aydogdu, A., Azzopardi, J., Bolzon, G., Bruschi, A., Drago, A., Garau, T., Gatti, J., Gertman, I., Goldman, R., Hayes, D., Korres, G., Lorente, P., Malacic, V., Mantziafou, A., Nardone, G., Olita, A., Ozsoy, E., Pairaud, I., Pensieri, S., Perivoliotis, L., Petelin, B., Ravaioli, M., Renault, L., Sofianos, S., Sotillo, M. G., Teruzzi, A., and Zodiatis, G.: Operational evaluation of the Mediterranean Monitoring and Forecasting Centre products: implementation and results, Ocean Sci. Discuss., 9, 1813-1851, https://doi.org/10.5194/osd-9-1813-2012, 2012.

Tonani, M., Salon, S., Korres, G., Bolzon, G., Clementi, E., Cossarini, G., Crise, A., Drudi, M., Fratianni, C., Girardi, G., Guarnieri, A., Marino, S., Oddo, P., Pinardi, N., Simoncelli, S., Solidoro, S., and Teruzzi, A.: The Mediterranean Monitoring and Forecasting Centre, a component of the MyOcean System, EGU General Assembly 2013, 7-12 April, 2013, Vienna, Austria, p.13628, 2013.

Vichi, M., Pinardi, N., and Masina, S.: A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: theory, J. Marine Syst., 64, 89-109, https://doi.org/10.1016/j.jmarsys.2006.03.006, 2007a.

Vichi, M., Masina, S., and Navarra, A.: A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part II: numerical simulations, J. Marine Syst., 64, 110-134, https://doi.org/10.1016/j.jmarsys.2006.03.014, 2007b.

Vichi M., Lovato T., Lazzari P., Cossarini G., Gutierrez E., Mattia G., Masina S., McKiver W. J., Pinardi N., Solidoro C., Tedesco L., and Zavatarelli M.: The Biogeochemical Flux Model (BFM): Equation Description and User Manual. BFM version 5.1. BFM Report series N. 1, Release 1.1, July 2015, Bologna, Italy, http://bfm-community.eu, pp. 104, 2015.

Volpe, G., Santoleri, R., Vellucci, V., Ribera d'Alcala, M., Marullo, S., and D'Ortenzio, F.: The colour of the Mediterranean Sea: Global versus regional bio-optical algorithms evaluation and implication for satellite chlorophyll estimates, Remote Sens. Environ., 107, 625–638, https://doi.org/10.1016/j.rse.2006.10.017, 2007.

Volpe, G., Colella, S., Forneris, V., Tronconi, C., and Santoleri, R.: The Mediterranean Ocean Colour Observing System – system development and product validation, Ocean Sci., 8, 869–883, https://doi.org/10.5194/os-8-869-2012, 2012.

Volpe, G., Pitarch, J., Colella, S., Brando, V.E., Forneris, V., Bracaglia, M., and Benincasa, M.: Quality Information Document for the OCTAC Products - Ocean Colour Mediterranean and Black Sea Observation Product. Copernicus Monitoring Environment Marine Service, 2017.

von Schuckmann, K., Le Traon, P.-Y., Alvarez-Fanjul, E., Axell, L., Balmaseda, M., Breivik, L.-A., Brewin, R.J.W., Bricaud,

5 C., and other 70 co-authors: The Copernicus Marine Environment Monitoring Service Ocean State Report, J. Oper. Oceanogr., 9, s235-s320, https://doi.org/10.1080/1755876X.2016.1273446, 2016.

Xing, X., Morel, A., Claustre, H., Antoine, D., D'Ortenzio, F., and Poteau, A.: Combined processing and mutual interpretation of radiometry and fluorometry from autonomous profiling Bio-Argo floats: 2. Colored dissolved organic matter absorption retrieval, J. Geophys. Res.-Oceans, 117, C04022, https://doi.org/10.1029/2011JC007632, 2012.

10 Weiss, R.F.: The solubility of nitrogen, oxygen and argon in water and seawater. Deep-Sea Research 17, 721-735, 1970.

Zacharioudaki,   A.,   Ravdas,   M.,   and   Korres,   G.:   Quality   Information   Document   for MEDSEA_ANALYSIS_FORECAST_WAV_006_017, Copernicus Marine Environment Monitoring Service, available at: http://cmems-resources.cls.fr/documents/QUID/CMEMS-MED-QUID-006-017.pdf, 2018.

Yu, L., Fennel, K., Bertino, L., El Gharamti, M., and Thompson, K. R.: Insights on multivariate updates of physical and

15 biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling, Ocean Model., 126, 13-28, https://doi.org/10.1016/j.ocemod.2018.04.005, 2018.

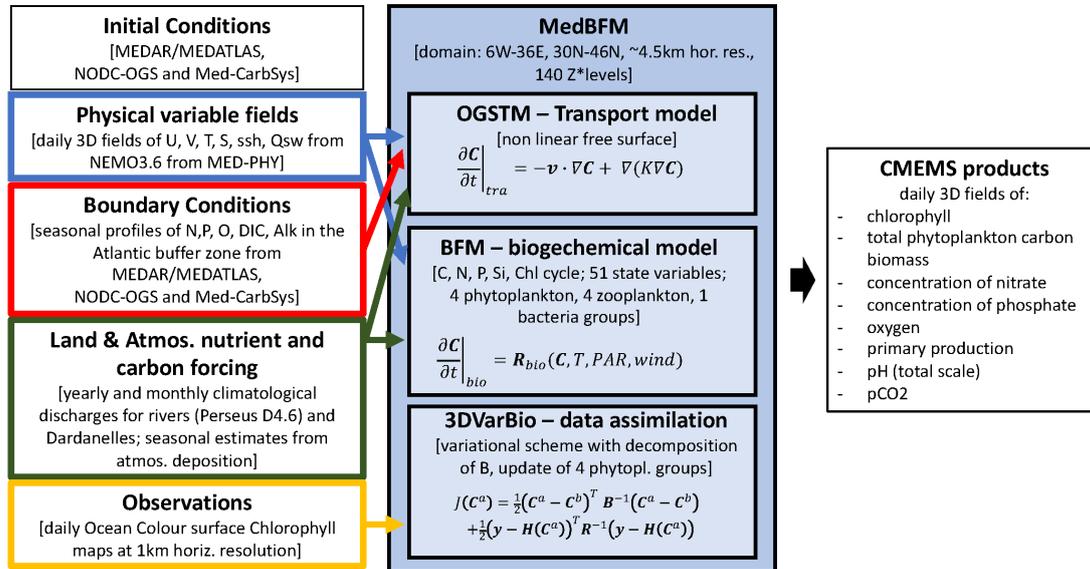



Figure 1: The MedBFMv2.1 model system and interfaces with other components of CMEMS and external forcing data.

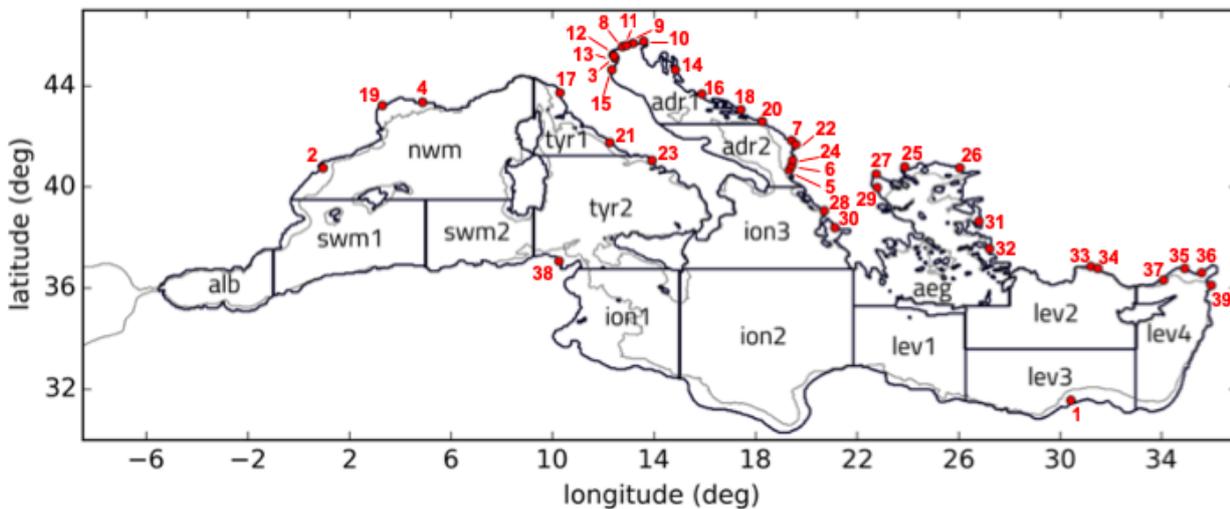



Figure 2: Subdivision of the model domain in sub-basins used for the validation of the qualification run. According to data availability and to ensure consistency and robustness of the metrics, different subsets of the sub-basins or some combinations among them can be used for the different metrics: lev = lev1+lev2+lev3+lev4; ion = ion1+ion2+ion3; tyr = tyr1+tyr2; adr = adr1+adr2; swm = swm1+swm2. The grey line defines the bathymetric contour at 200 m. Red dots with numbers correspond to river mouths positions: Nile (1), Ebro (2), Po (3), Rhone (4), Vjosë (5), Seman (6), Buna/Bojana (7), Piave (8), Tagliamento (9), Soca/Isonzo (10), Livenza (11), Brenta-Bacchiglione (12), Adige (13), Lika (14), Reno (15), Krka (16), Arno (17), Nerveta (18), Aude (19), Trebisjnica (20), Tevere (21), Mati (22), Volturno (23), Shkumbini (24), Struma/Strymonas (25), Meric/Evros/Maritsa (26), Axios/Vadar (27), Arachtos (28), Pinios (29), Acheloos (30), Gediz (31), Buyuk Menderes (32), Kopru (33), Manavgat (34), Seyhan (35), Ceyhan (36), Gosku (37), Medjerda (38), Asi/Orontes (39).

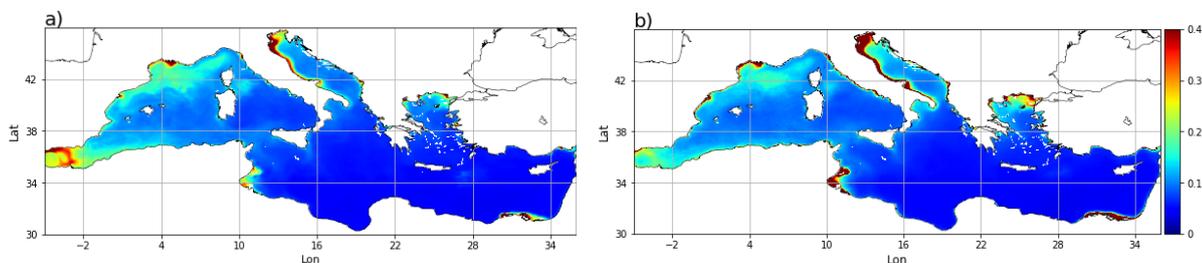
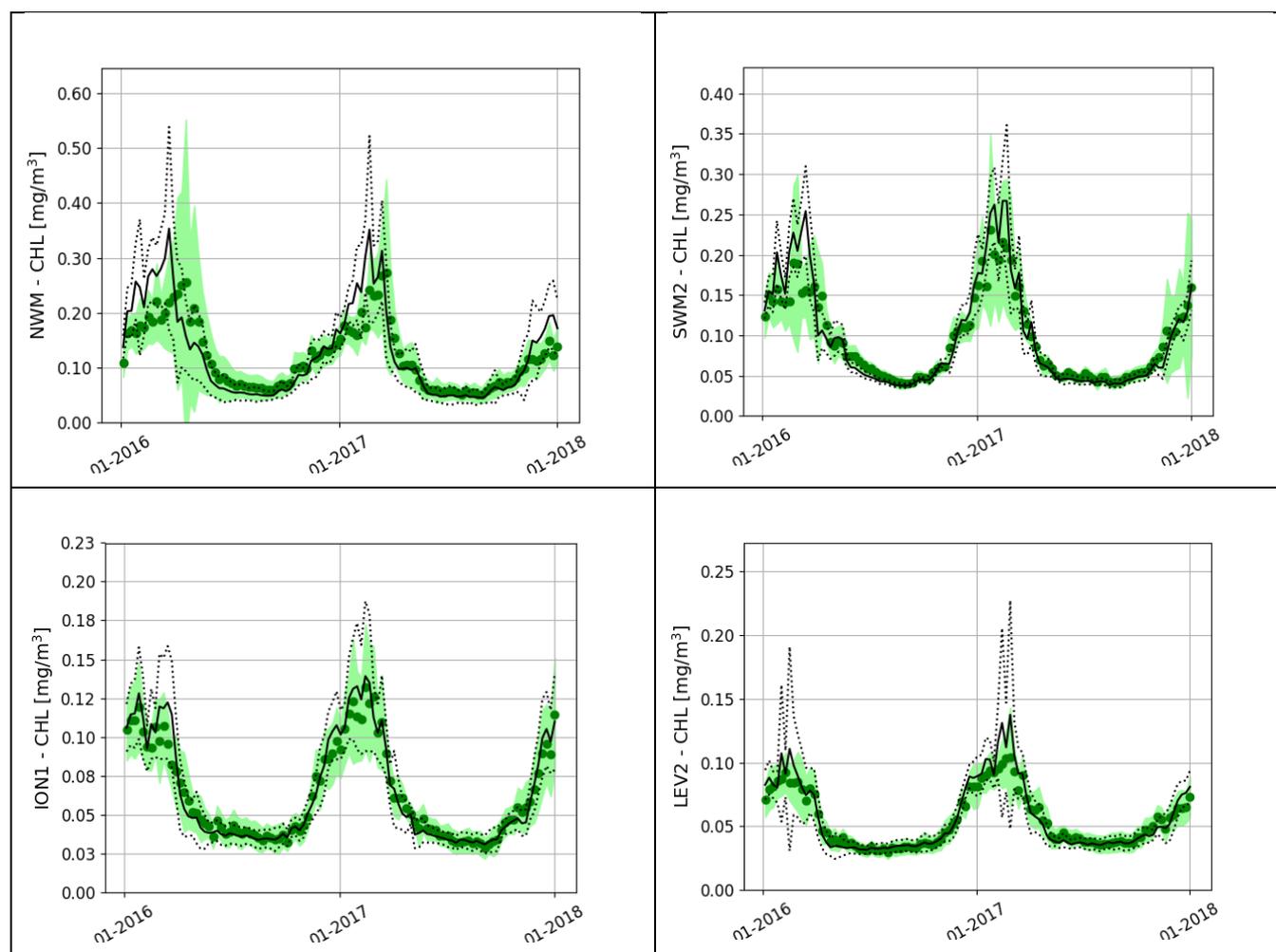



Figure 3: Averaged annual maps of surface chlorophyll (mg/m3) from 2016-2017 qualification run (a) and from NRT multi-sensor satellite (b).

5 **Figure 4: Model (black line, with standard deviation in black dots) and satellite (green dots, with standard deviation covering the light green area) time series of mean surface chlorophyll concentration in open sea areas in four selected sub-basins of Fig. 2.**

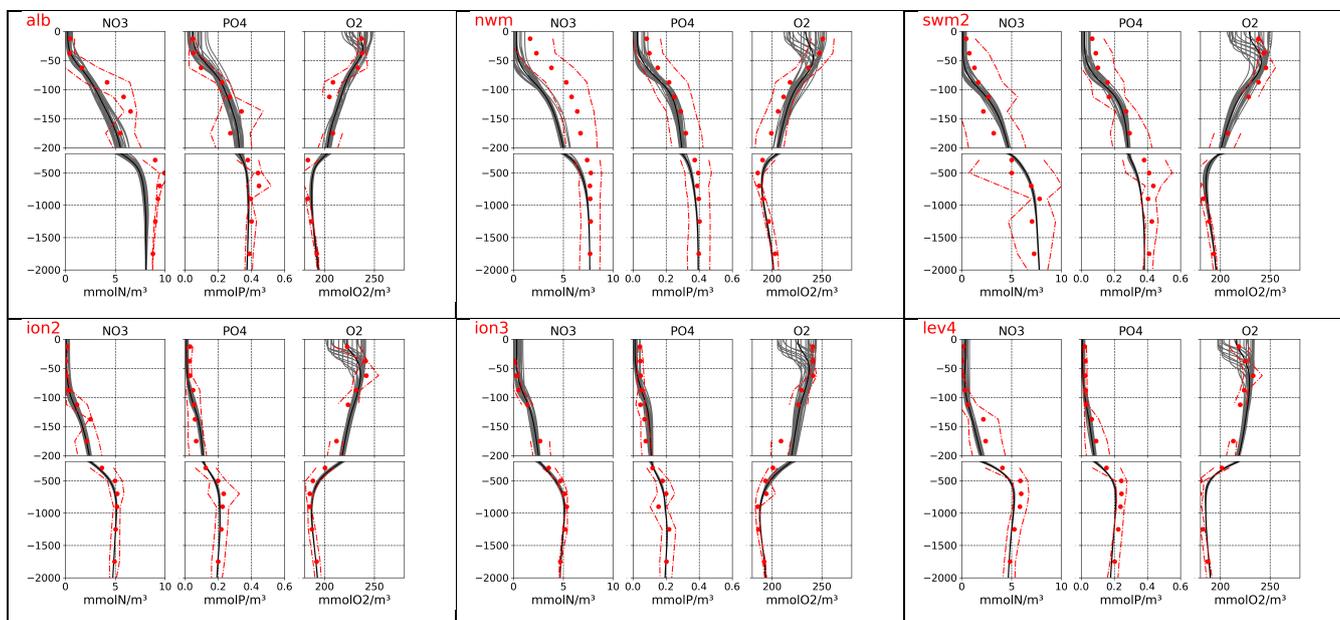
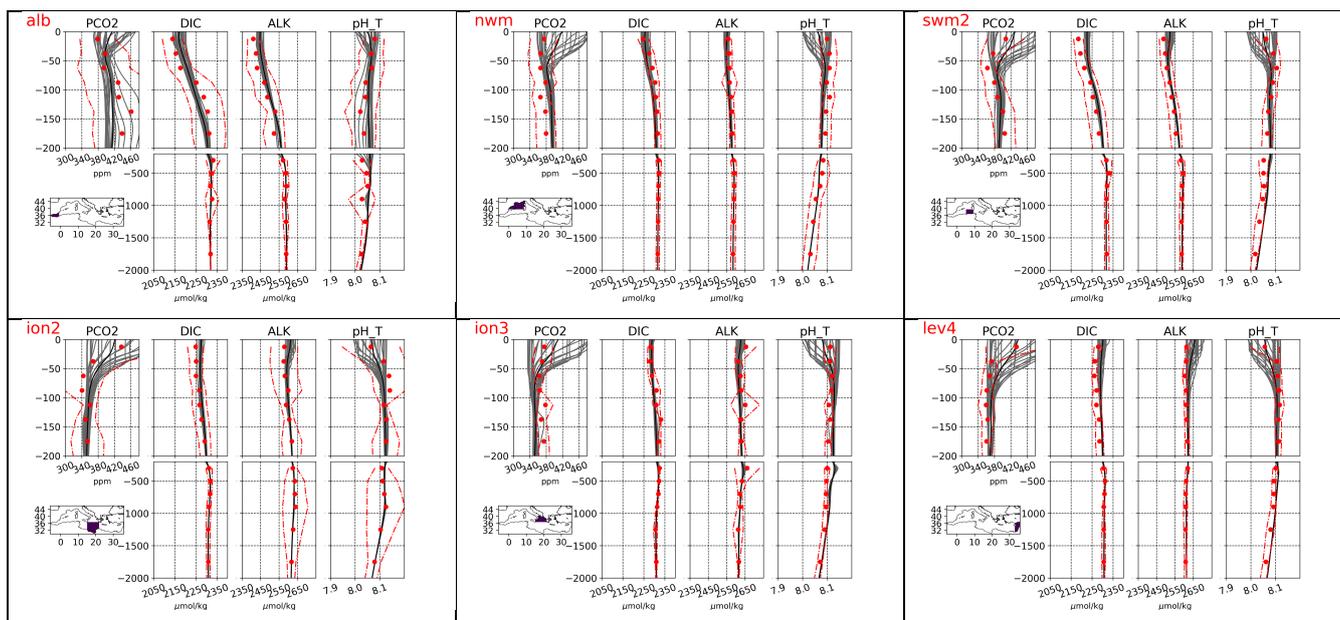



Figure 5: Monthly (grey lines) and mean (black lines) vertical profiles from the qualification run for selected sub-basins of Fig. 2 compared with climatological profiles (red dots) and variability ranges (one standard deviation, red lines) of nitrate, phosphate and dissolved oxygen retrieved from NODC-OGS dataset.

5   **Figure 6: Monthly (grey lines) and mean (black lines) vertical profiles from the qualification run for selected sub-basins of Fig. 2 compared with climatological profiles (red dots) and variability ranges (one standard deviation, red lines) of Dissolved Inorganic Carbon (DIC), alkalinity (ALK), pH in total scale and in situ condition (pH_T) and carbon dioxide partial pressure (pCO2). Climatological data for pH_T and pCO2 are reconstructed using CO2Sys software (Lewis and Wallace, 1998).**

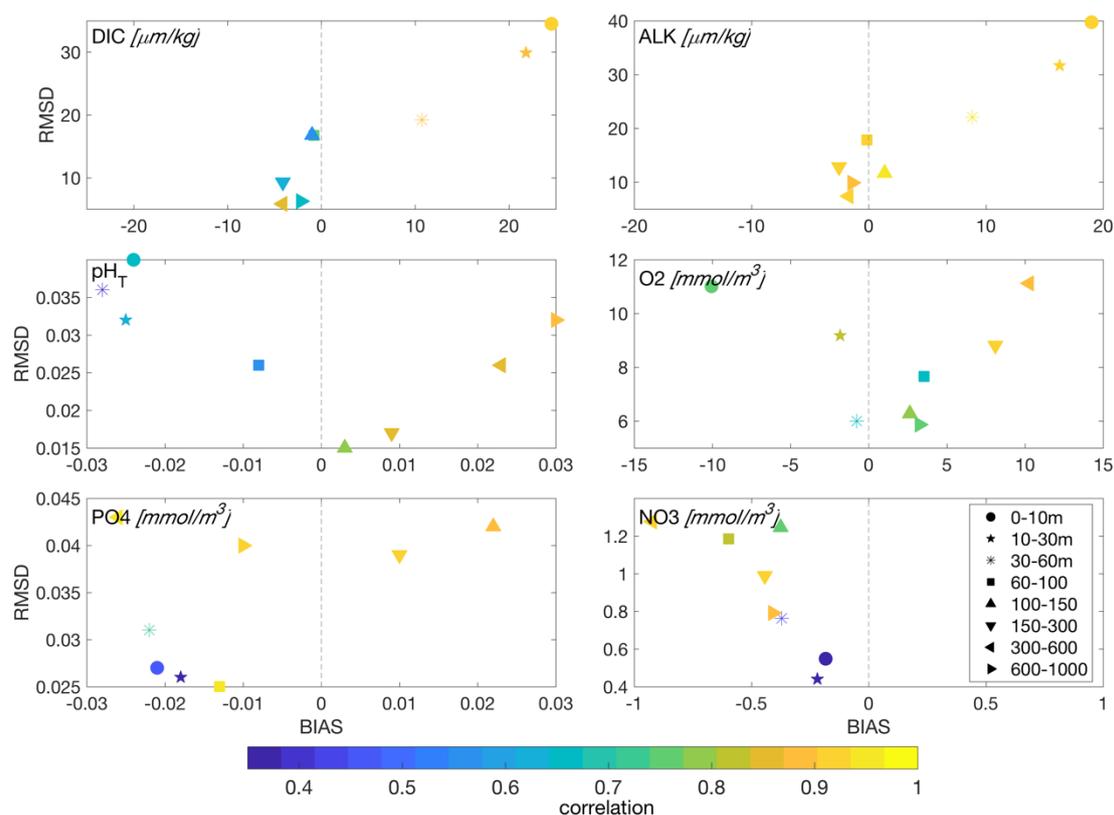



Figure 7: Target diagrams and correlation (values in colour shading) between model and climatology for the different layers (symbols) and for the selected variables: alkalinity (ALK), dissolved inorganic carbon (DIC), oxygen (OXY), phosphate (PO4), nitrate (NO3) and pH reported in total scale (pH_T).

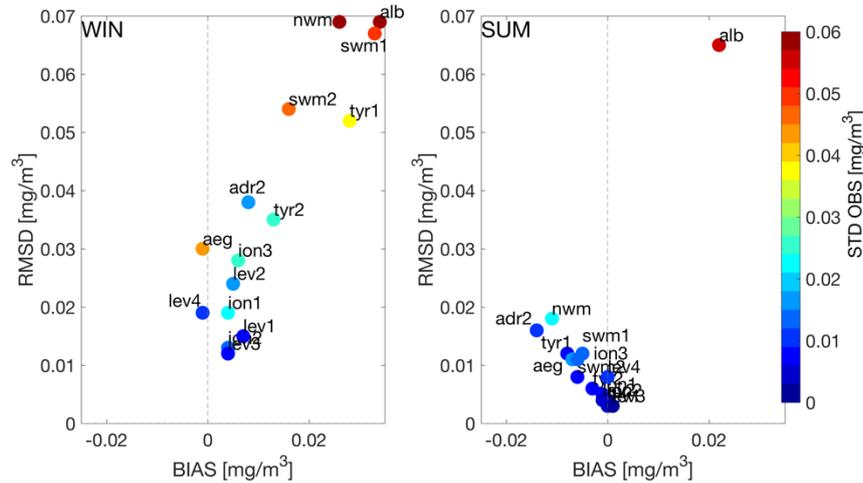



Figure 8: Target diagrams of the model and satellite chlorophyll comparison and standard deviation of observations (in colour shading) for two periods: January to April (WIN, left) and June to September (SUM, right). For sake of readability, an offset in the values in WIN of [RMSD, BIAS] for ALB and of [RMSD] for NWM (respectively equal to [0.17, 0.09] mg/m3, and [0.1] mg/m3) has been applied to include the dots within the plot.

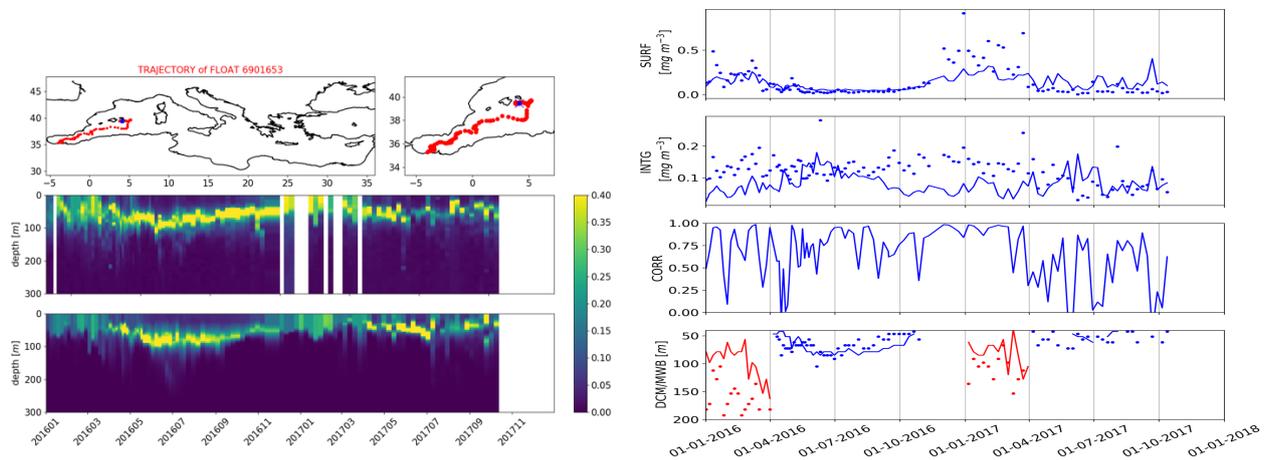



Figure 9: Time evolution of BGC-Argo float 6901653. Left panels: Hovmoller diagrams of chlorophyll concentration (mg/m³) from float data (2nd row) and model outputs (3rd row) matched-up with float position (1st row) for the period 2016-2017. Right panels: computation of selected skill indexes (1st to 4th row) for model (solid line) and float data (dots). The skill indexes are: surface (SURF) and 0-200m vertically averaged (INTG) chlorophyll, correlation (CORR), depth of the deep chlorophyll maximum (DCM, blue) and depth of the mixed layer bloom in winter (MWB, red). Trajectory of the BGC-Argo float is reported in the upper left panels (red dots), with deployment position (blue cross).

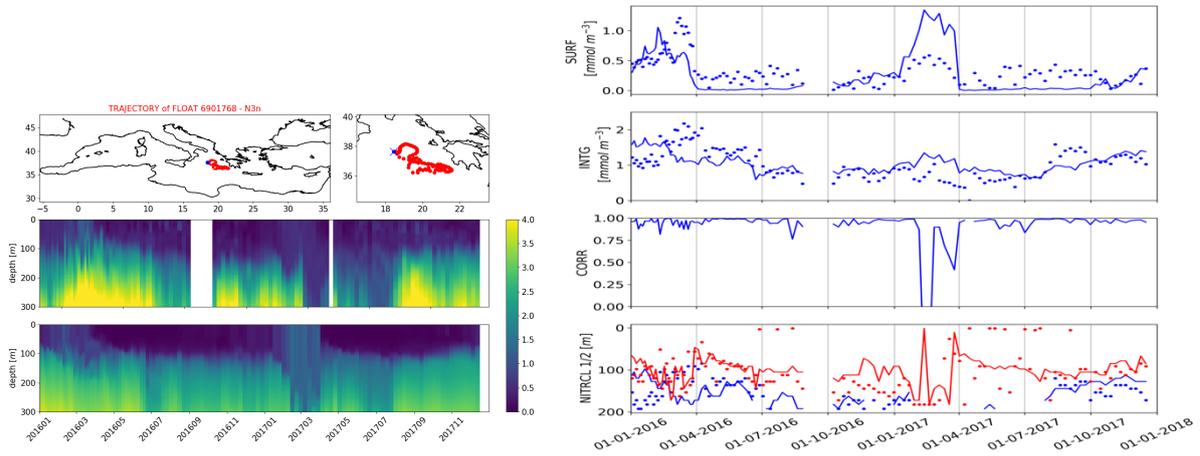



Figure 10: Time evolution of BGC-Argo float 6901768. Left panels: Hovmoller diagrams of nitrate concentration (mmol/m³) from float data (2nd row) and model outputs (3rd row) matched-up with float position (1st row) for the period 2016-2017. Right panels: computation of selected skill indexes (1st to 4th row) for model (solid line) and float data (dots). The skill indexes are: nitrate concentration at surface (SURF) and 0-200 m vertically averaged concentration (INTG), correlation between profiles (CORR), depth of the nitracline computed as NITRCL1 and NITRCL2. Trajectory of the BGC-Argo floats is reported in the upper left panels, with deployment position (blue cross).

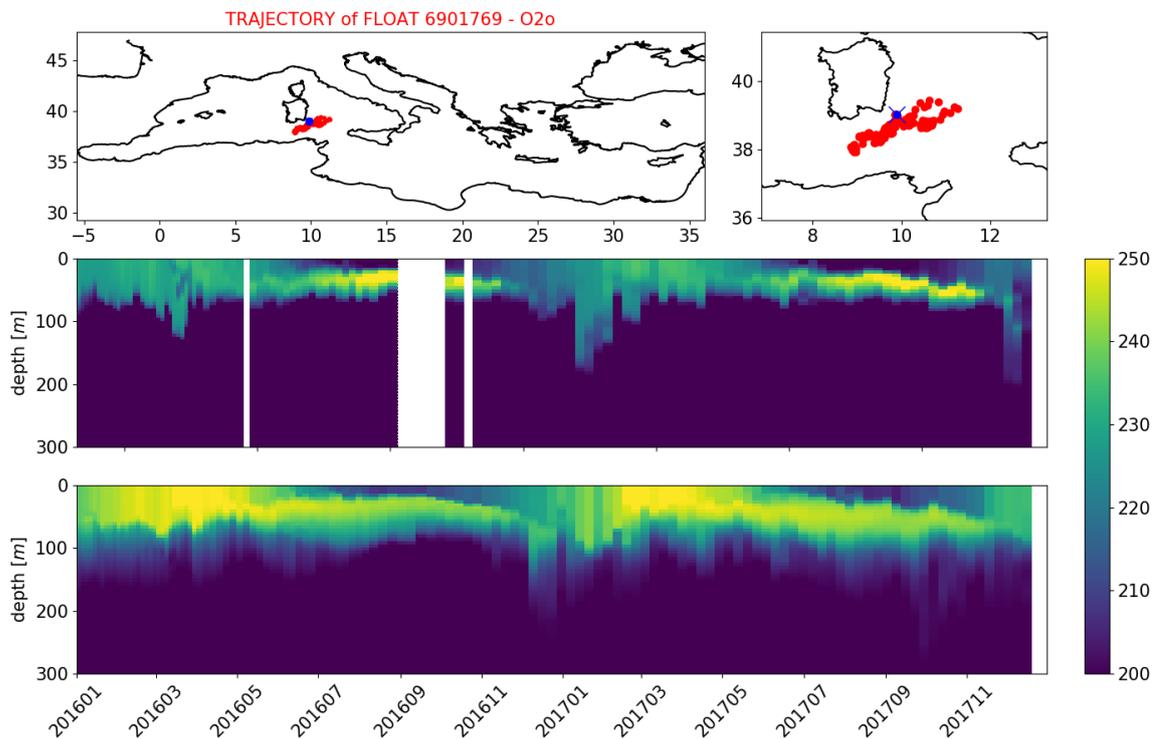



Figure 11: Hovmoller diagrams of oxygen concentration (mmol/m³) of one selected BGC-Argo float 6901769 (top panels) and model outputs (bottom panels) matched-up with float position for the period 2016-2017.

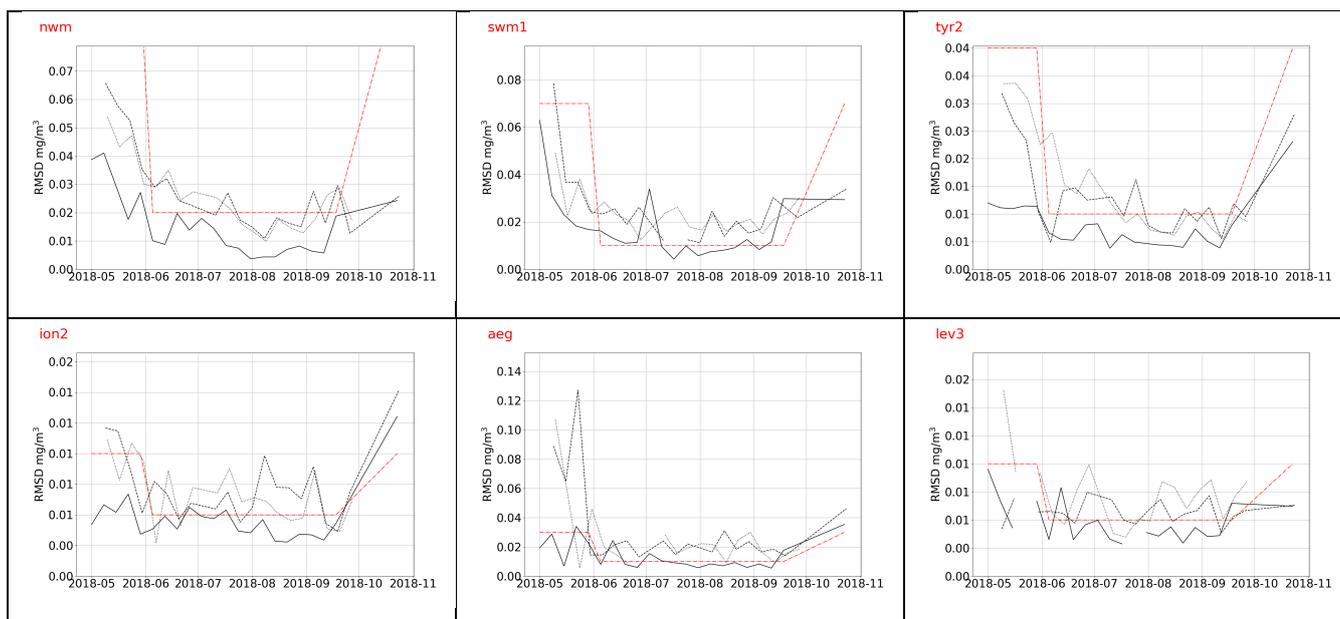



Figure 12: Sub-basin RMSD between surface chlorophyll model forecast at lead time 24 (black solid line), 48 (black dash line) and 72 hours (dotted line) and daily satellite maps. As benchmark reference, the 2016-2017 pre-operational mean RMSD are shown for the investigated period (red dashed line).

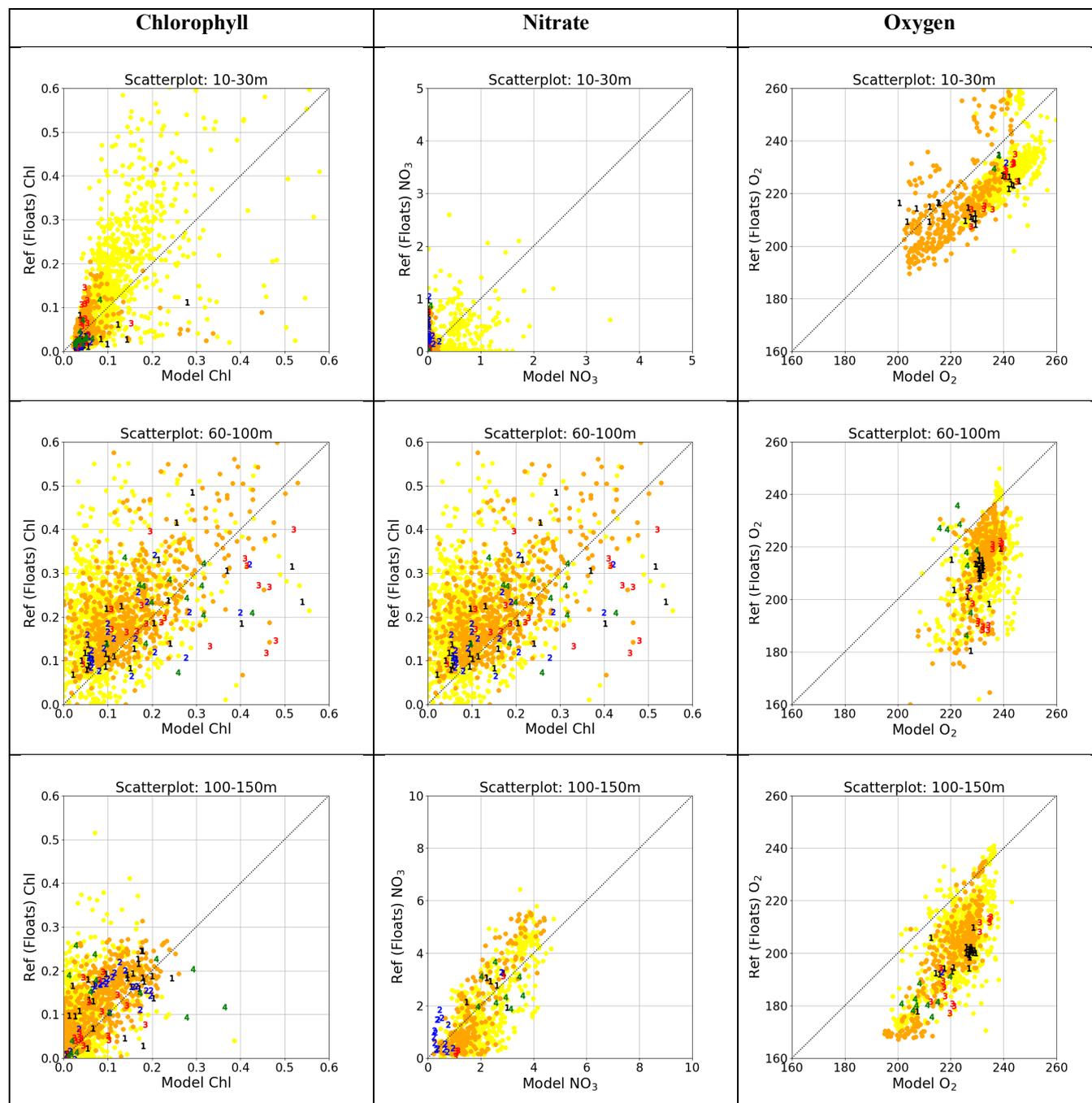



Figure 13: Scatter plots of reference (y-axis) versus model forecast (x-axis) for chlorophyll (left column), nitrate (middle column) and oxygen (right column) at different vertical layers: 10-30 m, 60-100 m and 100-150 m. Model forecast are labelled with numbers from 1 to 4 corresponding to lead time from T1 to T4. As benchmark reference, the 2016-2017 pre-operational results are shown for the period of investigation (May to August, orange dots) and for the other periods (yellow dots).

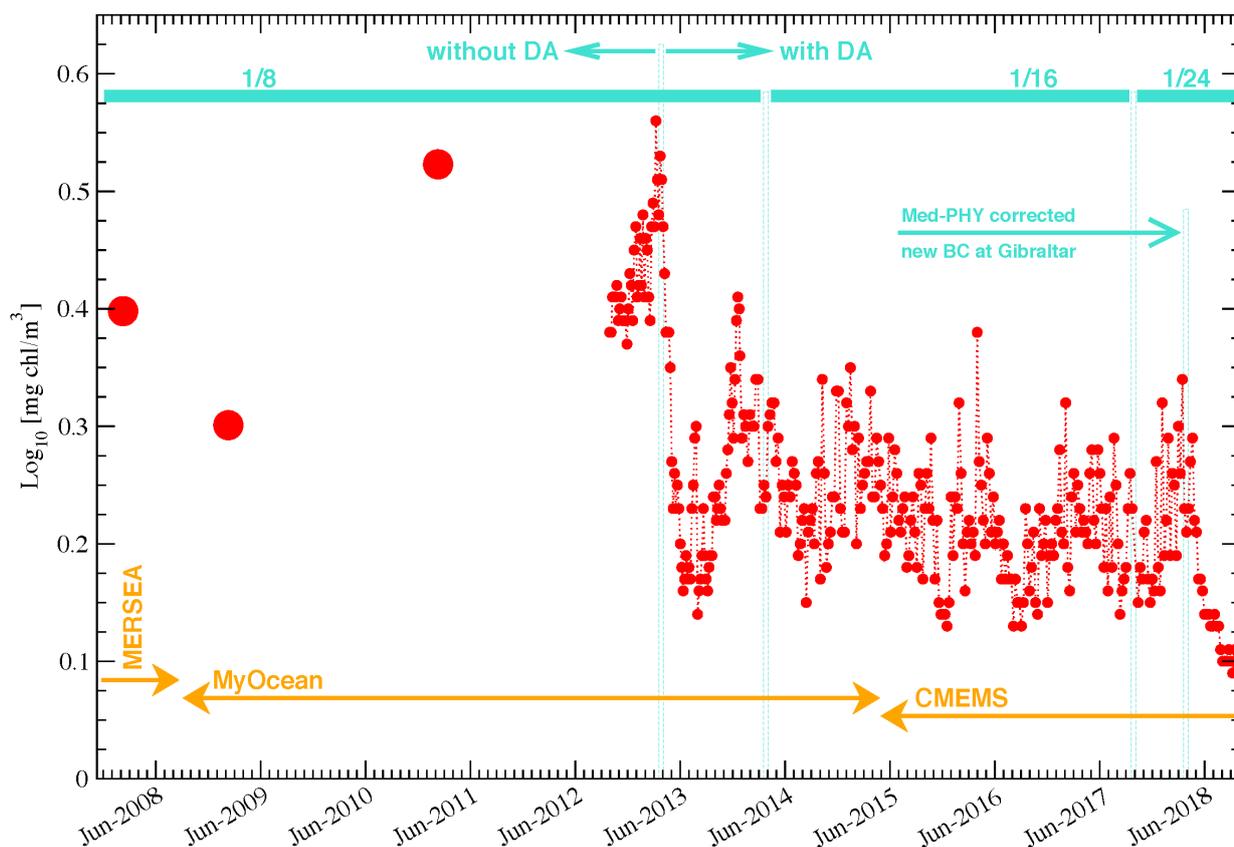



Figure 14: RMSD of the surface chlorophyll concentration with satellite data. The use of logarithmic units has been the standard for RMSD since the implementation phase. Regular, weekly product quality assessment (red dots and dashed line) started at the end of 2012, in the previous period quality assessment was performed only occasionally (e.g. Teruzzi et al., 2011; Tonani et al., 2012). In the plot, we identify the different projects (yellow), the start of data assimilation (April 2013), and the increment of horizontal resolution (azure).

	n. of BGC-Argo floats active	n. of profiles
Chlorophyll	28	2532
Nitrate	13	1406
Oxygen	15	1596

Table 1: Synthesis of the BGC-Argo floats dataset for Mediterranean Sea used in the present study, for chlorophyll, nitrate and oxygen.

Annual mean [gC/m²/y]	MODEL Lazzari et al. (2012)	SATELLITE Bosc et al. (2004)	SATELLITE Colella (2006);	In situ Siokou-Frangou et al., 2010	CMEMS qualification run
Mediterranean Sea (MED)	98±82	135.5	90±48		127±42
Alboran Sea (ALB)	274±155	230	179±116		249±56
South West Med –West (SWM1)	160±89	162	113±43		188±22
South West Med –East (SWM2)	118±70	162	102±38		162±12
North West Med (NWM)	116±79	170	115±67	105.8-119.6; 86-232*; 140-170**	149±18
Levantine (LEV1+LEV2+LEV3+LEV4)	76±61	105	72±21	59***	105±40
Ionian Sea (ION1+ION2+ION3)	77±58	120	79±23	61.8	107±18
Tyrrhenian Sea (TYR1 + TYR2)	92±5	137	90±35		139±25

5 **Table 2: Annual averages and short period estimates of the vertically integrated primary production for some selected sub-regions. Estimates are from multi-annual simulation (Lazzari et al., 2012), from satellite model (Bosc et al., 2004; Colella, 2006), from in-situ estimates (Siokou-Frangou et al., 2010) and from the present CMEMS qualification run. Notes: *only DYFAMED station; **Southern Gulf of Lions; ***only Cretan Sea.**

	SURF [mg/m³]		INTG [mg/m³]			DCM [m]		MWB [m]		average number of available profiles per month
	BIAS	RMSD	BIAS	RMSD	CORR	BIAS	RMSD	BIAS	RMSD	
alb	-0.04	0.24	-0.01	0.05	0.48	-19	19	n.c.	n.c.	2
swm	-0.02	0.07	-0.07	0.08	0.70	4	15	n.c.	n.c.	4
nwm	-0.07	0.18	-0.05	0.07	0.78	-9	15	43	61	12
tyr	-0.04	0.07	-0.07	0.08	0.72	-5	13	n.c.	n.c.	10
adr	0.01	0.04	-0.04	0.05	0.71	-3	12	n.c.	n.c.	4
ion	-0.03	0.06	-0.03	0.04	0.85	7	14	-29	56	21
lev	-0.01	0.05	-0.04	0.05	0.73	4	18	-52	60	22

10 **Table 3: Averages of the chlorophyll indicators based on the BGC-Argo floats and model comparison for the period January 2016 – December 2017. The indicators are the BIAS and RMSD of the surface (SURF) and of the vertically 0-200 m averaged (INTG) chlorophyll concentration, the correlation between model and BGC-Argo float data (CORR), the BIAS and RMSD of the depth of the Deep Chlorophyll Maximum (DCM) and depth of the vertically mixed winter bloom (MWB). Statistics are computed for selected aggregated sub-basins; MWB statistics are not computed (n.c.) for some sub-basins.**

	SURF [mmol/m³]		INTG [mmol/m³]		CORR	NITRCL1 / NITRCL2 Mean OBS [m]	NITRCL1 [m]		NITRCL2 [m]		average n. of available profiles per month
	BIAS	RMSD	BIAS	RMSD			BIAS	RMSD	BIAS	RMSD	
swm	0.13	0.22	0.04	0.78	0.92	100 / 106	-7	27	-23	38	1
nwm	-0.12	0.53	-0.61	1.44	0.98	64 / 96	35	39	2	20	2
tyr	-0.01	0.18	-0.68	0.76	0.97	82 / 82	8	17	-5	19	4
ion	-0.04	0.29	-0.03	0.36	0.93	148 / 111	-1	38	-14	37	6
lev	0.18	0.26	0.53	0.63	0.93	176 / 153	-31	37	-42	57	17

Table 4: Averages of the nitrate indicators based on the BGC-Argo floats and model comparison for the period January 2016 – December 2017. The indicators are the correlation between model and BGC-Argo float data (CORR), the BIAS and RMSD of the surface (SURF) and of the vertically 0-200 m averaged (INTG) nitrate concentration, the BIAS and RMSD of the depth of the nitracline computed as NITRCL1 and NITRCL2. Statistics are computed for selected aggregated sub-basins. For a reference, the mean value of NITRCL1 and NITRCL2 estimated from BGC-Argo data is included (Mean OBS).

	0-10 m	10-30 m	30-60 m	60-100 m	100-150 m	150-300 m	300-600 m	600-1000 m	average n. of available profiles
swm	13.2	13.1	12.4	27.5	27.2	27.3	25.1	18.8	26
nwm	14.6	14.7	19.0	25.6	25.0	24.3	23.8	20.6	107
tyr	15.3	14.4	14.8	24.5	24.2	24.5	25.5	21.5	217
adr	5.8	12.2	8.1	5.0	3.9	3.7	4.5	10.4	78
ion	12.9	11.9	10.2	14.4	17.9	18.4	18.6	16.6	242
lev	13.9	12.5	11.9	18.0	21.3	21.2	25.2	21.7	388

Table 5: RMSD of the oxygen difference between BGC-Argo float and model at the float position and time. Statistics are computed for sub-basins and given layers.

	PRE-OPERATIONAL	T0	T1	T2	T3
NO3	0.78	0.79	0.68	0.63	0.60
O2	18.05	20.08	18.62	25.09	22.46
CHLA	0.13	0.06	0.05	0.07	0.08

Table 6: RMSD of BGC-Argo float and model comparison for the pre-operational run and for the first 4 days of forecast of the Med-BIO forecast system (T0 to T3) since April 2018. Statistics are computed using the layers 0-300 m for nitrate and oxygen and 0-150 m for chlorophyll.