Dear Editor,
please find here the revision of our manuscript with point-to-point responses to the comments made by the two Reviewers. We wish to thank the Reviewers for the insightful comments that have contributed to improve our manuscript.
In particular:

- The discussion on how the results extracted from metrics application are related with the possible causes of uncertainty (e.g. boundary conditions, physical forcing, accuracy of the reference datasets) has been revised and extended (R2.M2, R1.2, R1.3 and R1.4). The title has been changed from "Marine Ecosystem forecasts: skill performance of the CMEMS Mediterranean Sea model system" to "Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts" (R1.5).
- The definition of the novel metrics used for the model - BGC-Argo comparison have been better presented and discussed (R1.1).
- References to other biogeochemical modelling applications in Mediterranean Sea, and their specific validation framework, have been added (R2.M1).
- Quality of Figures 9, 10 and 12 has been improved and, when feasible, the figures concerning the NRT validation have been updated with the most recent data (R2.9-12 and R2.15).
- Abstract has been modified highlighting some results which were not mentioned in the submitted version (quality improvement along the past decade, the importance of data availability to build robust statistics, the NRT vs seasonal benchmark comparison as a monitoring tool of the operational system).
- Bibliography references have been reviewed.
- Relevant language editing has been performed, wording inconsistencies (as "data set" / "dataset") have been solved, and typos have been corrected (several comments by R1 and R2).

Further, we carefully considered all the other technical points raised by the reviewers.

The points raised by the Reviewers are in *BLACK ITALIC*, our responses in BLUE and the proposed modifications to the manuscript in *GREEN ITALIC*, with line numbers of the revised version. To easily refer through the different reviews, we added a reference to each point raised by the two Reviewers, labelled as "Rx.y", where "x" is the n. of Reviewer, and "y" is a number when referred to a specific comment, or a letter when referred to a technical correction (for the 2 general comments of Reviewer #2, we use R2.M1 and R2.M2). For sake of clarity, when in our replies we refer to specific sentences in the revised manuscript, we use "(RPn,Lm-k)", where "n" is n. of page and "m-k" are specific lines.

Attached to this letter, after the point-by-point responses, we included the revised version of the manuscript with track changes.

**Reply to comments from the Reviewers**

Reviewer #1 – Specific comments

*R1.1. (P8 L1-10) Authors discuss further in the manuscript that direct comparison of model results with the sensor data should be done with caution as the applied corrections may not reflect true value. Thus, application of the introduced metrics which is consistent within itself (e.g. normalising* *the data to its own surface value) is a good approach. As these are new metrics, for their applicability to other regions, the choice of the criteria deserves a discussion. Do 10% of the surface chl value, or 2mmol/m3 nitrate have a significant meaning? Are these values applicable enough throughout the Mediterranean or have the authors seen regional inconsistencies? Are the strict choice of seasons (jan-mar and apr-oct) valid in practice (can't tell much from Figure 9, but wouldn't using MLDs from the ARGOs yield better estimation of which criteria to use? ARGO profiles show deep chl formation in late Jan and late March 2016 hence much deeper mwb than then model)*

REPLY - The metrics concerning chlorophyll and the relative two periods for the computation are based on published phenomenological understanding. In particular, they have been derived from the outcomes of Lavigne et al. (2015), who identified some standard shapes for chlorophyll profiles from the analysis of a large number of fluorescence data in the Mediterranean Sea (see their Fig. 2). In particular, our summer period is based on the consideration that the DCM profile shape (useful to define our DCM depth index) is typically observed from April to October (Fig. 5 of Lavigne et al., 2015). Then, three other profile shapes (i.e., "homogeneous", "high surface chlorophyll-HSC" and "complex" in Fig. 2 of Lavigne et al., 2015) are characterized by steady-decreasing values with depth, and typically occur between January and March in different Mediterranean regions (Fig. 5 of Lavigne et al., 2015).
The choice of the 10% criterion for the MWB depth index was made after a sensitivity analysis varying the limit between 1 to 10% (not shown). The value of 10% gave results qualitatively consistent with those reported by Lavigne et al. (2015). Further, lower percentage values gave more unclear patterns because the depth of the MWB increases substantially and the thickness of the model layers has also an impact, due to the vertical resolution coarsening (i.e. thickness of model layers is 5 m at 70 m depth, 7 m at 100 m and 10 m at 140 m).
Alternatively, a statistical analysis and additional data (such as Mixed Layer Depth - MLD - values, as suggested by the Reviewer) can provide a better tool to identify the profile shapes and their distribution in time. In particular, the use of the MLD for the choice of seasons' limits surely represents a promising alternative option, since it would account for the specific physical conditions at each float profile, and beside that, it might be useful to catch possible errors propagated by the physical forcing to biogeochemical processes (e.g., comparing model-derived MLD with float-derived MLD). However, it must be noted that there are several possible definitions of MLD which can give slightly different results, thus a sensitivity analysis of the biogeochemical metrics with respect to the MLD definition would be necessary to tune the choice of the criteria based on MLD (this approach is under investigation, and will be developed within the CMEMS R&D activities). The implementation of such methodology might require a considerable amount of data and is out the scope the present work. Therefore, we decided to adopt an a priori, while rigid, temporal subdivision based on literature to test the feasibility of the computation of the novel chlorophyll metrics.

Regarding the criteria for nutricline metrics, they have been designed to track the time evolution of the nitrate vertical profile. Being aware that selecting an unique criterion to detect the nitracline might be sensitive and controversial, we proposed two different approaches: the depth of the 2 mmol/$m^3$ concentration isopleth (NITRCL1) and the depth of the maximum nitrate vertical gradient (NITRCL2). Following Manca et al. (2004), who showed that the values of nitrate concentration at depth higher than 400 m are around 4-5 in the eastern basin and 6-7 mmol/$m^3$ in the western, we considered the isopleth of 2 mmol/$m^3$ a safe value to detect the rapid change between the very low concentration typically measured at the surface and the high concentration at depth in all areas of the Mediterranean. As discussed for Fig. 10 (see R2.11), the two indexes were able to show different aspects of the nitrate profile evolution, justifying their use to provide indications aimed to monitor the model error behaviour.

We added more details concerning the MWB, DCM and NITRCL1/2 metrics at the end of Section 4.1 (RP8,L27-32 and RP9,L1-9):

*The definitions of DCM and MWB metrics are consistent with the outcomes of Lavigne et al. (2015), who identified some standard shapes for chlorophyll vertical profiles and their temporal distribution from the analysis of a large dataset of fluorescence data in the Mediterranean Sea (see their Figs. 2 and 5). In particular, the summer period defined to estimate the DCM index is based on the consideration that the DCM profile shape is typically observed from April to October. Otherwise, the choice to limit the estimate of the MWB index from January to March is motivated by the fact that steady depth-decreasing profiles typically occur during that period in different Mediterranean regions. Further, the choice of the 10% criterion for the MWB index was set after a sensitivity analysis varying the threshold between 1 to 10% (not shown), with the 10% value giving results qualitatively consistent with those reported by Lavigne et al. (2015).*
*The rationale behind the nitracline depth metrics is defining an index useful to track the time evolution of the nitrate profile. Being aware that the choice of a specific value of nitrate concentration may be controversial, we propose two different indexes: the first is based on the depth of the 2 mmol/m$^3$ concentration isopleth (NITRCL1), the second is related to the depth of the maximum nitrate vertical gradient (NITRCL2). According to Manca et al. (2004), the values of nitrate concentration at depth higher than 400 m are around 4-5 mmol/m$^3$ in the eastern basin and 6-7 mmol/m$^3$ in the western, therefore the 2 mmol/m$^3$ isopleth can be considered a consistent threshold to detect the rapid change between the very low concentration typically measured at the surface and the high concentration at depth in all areas of the Mediterranean Sea.*

*R1.2. (P11 L8-10) Does this partly explain the lower modelled NO3 concentrations (e.g. nwm) due to the lack of N river load time-series? How does model perform in terms of N/P ratios? Does it represent the high N/P ratio character of the Med Sea and its regional differences?*

REPLY - We agree: the effect of the lack of high frequency data of nutrient discharges is one of the most important sources of uncertainty at the daily/weekly time scale (not at seasonal/annual scale) and at very local coastal scales, as discussed by Teruzzi et al. (2018). Indeed, we highlighted this potential issue in the submitted manuscript, and reported in the revised one at (RP12,L13-15). Uncertainty in nwm is partly related to a possible underestimation of the river input forcing and possibly to the effect of lateral circulation from alb and swm1 surface waters (see Fig. 5). A sensitivity analysis of the impacts of the different factors would help in elucidating the most relevant factor. We added a sentence highlighting the specific case of nwm in Section 5.1.1, at (RP10,L13-15):

*Uncertainty in nwm upper layer nitrate (Fig. 5) is partly related to a possible underestimation of the Ebro/Aude/Rhone rivers input forcing and possibly to the effect of lateral circulation from Alboran Sea and Southern Western Mediterranean surface waters (see Fig. 5, panels “alb” and “swm2”).*

An important aspect to consider when discussing the model performance is that the fine subdivision of the Mediterranean Sea in 16 sub-basins allows to detect the relevant spatial gradients and, thus, to highlight possible issues, as discussed in the submitted version, and reported in the revised one at (RP17,L28-32 and RP18,L1-3). The choice of the 16 sub-basins, designed by the Med-BIO component of Med-MFC and agreed as common skill assessment protocol for the other two components (Med-PHY and Med-WAV), was a trade-off among having a number of areas as larger as possible, the need for having robust in situ statistics and the known characteristics of the dynamics derived from literature. Thus, our proposed subdivision is a relevant result in itself. We included this point in the Discussion on the sub-basin subdivision at (RP18,L3-5), adding a comment on the detection of the surface nitrate underestimation of Fig. 5:

*[…], justifying a posteriori our sensible definition of the 16 sub-basins. In fact, the comparison of nutrients profiles of Fig. 5 highlights the satisfactory model performance in reproducing the mean spatial gradients and the possible anomalies, such as the underestimation of upper layer nitrate in nwm sub-basin.*

Concerning N/P ratios, the performance of OGSTM-BFM model system was assessed in Lazzari et al. (2016) for nitrate and phosphate separately, showing a general higher-than-Redfield ratio in the Mediterranean Sea (closer to a N:P=22 ratio, with exception of the Alboran Sea area, characterized by a lower-than-Redfield ratio) and a significant spatial variability. The present operational configuration of the MedBFM incorporates the results presented by Lazzari et al. (2016; see their Fig. 7 here reported as Fig.R1, for further details please refer to the paper) and provides consistent results on N:P ratio (see also N and P values in Fig. 5).

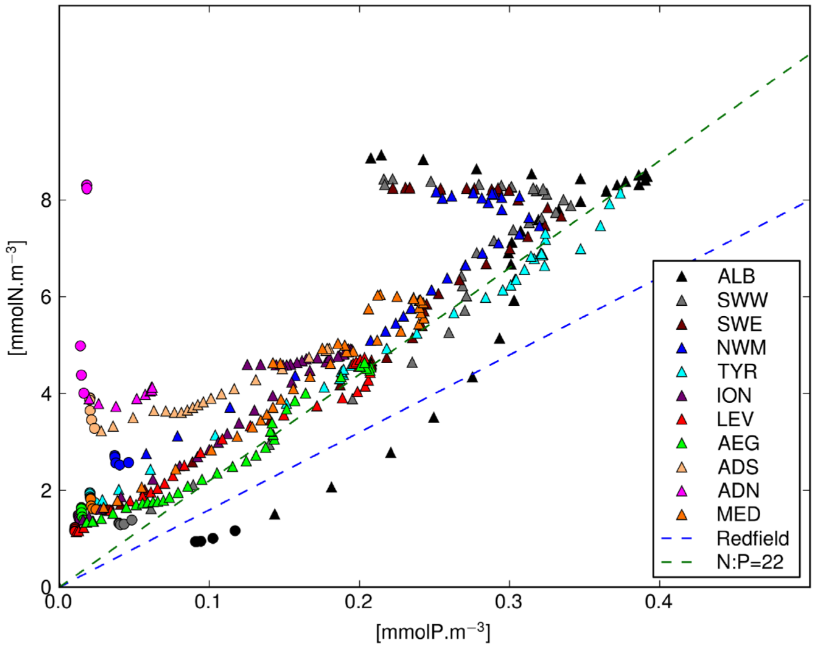

Fig.R1 - Results from Lazzari et al. (2016), model-derived vertical profiles of phosphate (x-axis) and nitrate (y-axis) averaged by sub-basins (reported in Fig. 1 of Lazzari et al., 2016) for the period 1999–2004. Each entry represents the spatial-temporal average at a certain depth: 0–50 m layer (circles) and 50-bottom layer (triangles).

*R1.3. (Fig9) How does BGC-Argo surface chl compare with the satellites? P11 L4 suggest the model has a higher (0.015) bias for the winter (model vs satellite), supported by Fig8 with more pronounced bias for the west/northwest Med, while paragraph of P11 L32 suggest the model has lower values when compared to BGC-Argo data. Is there a consistent ratio between satellite and float data, and how applicable is it to use global correction of division by 2 as suggested by Roesler et al. (2017) taking into account the regionally different ratios shown in the same article. As the Mignot et al. (2018) manuscript is in review, I cannot comment about their results but can the application of their method suggest different correction factors with a better regional fit?*

REPLY – We thank the Reviewer for having raised this point, since the inconsistency in surface chlorophyll observations between satellite and BGC-Argo floats has been already observed in our investigations (see Fig.R2, showing that positive deviations of BGC-Argo floats w.r.t. satellite data are evident in winter, while negative deviations of BGC-Argo data are present in summer and spring seasons) and represents a potential issue, not only for validation purposes as discussed here, but also for multi-platform data assimilation (e.g., combined assimilation of chlorophyll from BGC-Argo floats and satellite). Regional corrections of BGC-Argo floats data can be advisable (as shown in Roesler et al., 2017), such as regional algorithms for the ocean color exist for satellite, to provide chlorophyll in different regions of the global ocean. However, at the present stage, as shown by Roesler et al. (2017), regional calibration factors appear to be based on very limited statistics: for this reason, we think it is more robust to rely on the recommend factor of 2 global bias correction.

Methods such as that of Mignot et al. (2019) can be helpful to give a better estimate of the observational error (of great importance in data assimilation), proven the availability of a sufficient amount of in situ independent data.

Such investigation is off-topic with respect to our manuscript, but emphasizes that operational systems are optimal tools to test the consistency of many different sources of information, as was already reported in the submitted manuscript citing She et al. (2016), and in the revised version at (RP2,L25-26) and (RP18,L25-26).

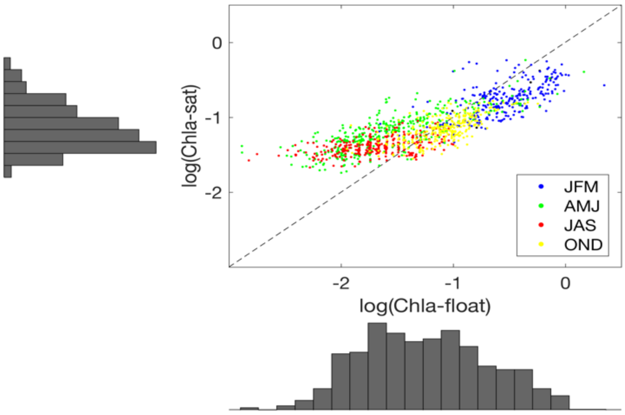

Fig. R2 - Chlorophyll concentration comparison between surface BGC-Argo floats data and satellite data (CMEMS ESA-CCI Mediterranean data). Data of all floats available between 2015-2017 are here included. The matchup between data is performed on a daily basis and using a bilinear spatial interpolation.

*R1.4. (P14 L20-23) Authors point out that the introduced BGC-ARGO related metrics are already being implemented for data assimilation purposes with consequent improvements in model* *solutions. Before the assimilation phase (e.g. pre-opetaional runs), does the skill assessment documented here (of BGC-Argo metrics) reveal any prior messages for model parameter adjustment such as for light attenuation or nutrient assimilation rates, or errors of physical model origin? I can see the use of this dataset not only for forecasting purposes and skill assessment purposes, but its high resolution coverage including ocean interior is of high value. A short comment on that would be good scientific addition to the manuscript.*

REPLY – We thank the Reviewer for this comment. The integration of BGC-Argo floats within an analysis and forecasting system (in terms of data assimilation) paved the way to an in-deep study of the interior of the sea and its dynamics. Then, given the fact that the BGC-Argo also provide profiles of PAR, salinity and temperature, which are forcing mechanisms of the biogeochemical processes (or proxies for the forcing impacting the biogeochemistry), an analysis of the model uncertainty can be made using multivariate statistical analysis, with the aim of disentangling the sources of error on biogeochemical profiles. This point has been also reported in comment R2.M2 (e.g., issues related to inaccuracy of the boundary conditions, oxygen float data quality, model surface chlorophyll discrepancy).
Beside the validation, it is worth to note that specific physical (such as MLD and euphotic zone depth) and biogeochemical (such as NITRCL1/2, DCM and MWB) indexes that can put in relation the shape and intensity of the profiles with the underlying processes would allow to better investigate coupled vertical physical-biogeochemical processes. We think that our work provides a first step to identify and quantify several functional biogeochemical indexes. However, a critical point is the availability of a sufficient amount of profiles for variables like nitrate and oxygen, which may allow for statistically significant analysis.
As also discussed in R2.M2 reply, we included a short comment on this point in the Discussion section (RP18,L26-34):

*In perspective, the integration of BGC-Argo within operational ocean forecasting systems in terms of data assimilation (see Cossarini et al., 2019) becomes strategical for an in-deep study of the interior of the sea and its dynamics. Moreover, considering that BGC-Argo floats also provide profiles of physical quantities (i.e., radiometric quantities – PAR- and temperature), an analysis of specific physical (e.g., MLD, euphotic zone depth) and biogeochemical (e.g., NITRCL, MWB, DCM) indexes that can reveal relationships between the shape and/or intensity of the profiles and the underlying dynamics would allow to further delve into coupled vertical physical-biogeochemical processes. In such a view, our work provides a first step to identify and quantify several functional biogeochemical indexes. Nonetheless, a critical point remains the availability of a sufficient amount of profiles for variables like nitrate and oxygen, which may allow for statistically significant analysis.*

*R1.5. I see that the manuscript is designed as a document for the overall skill assessment of MedBFM, but both the abstract and the manuscript throughout have stressed the importance and usefulness of their new metrics (GODAE Class 1 and 4, and especially the use of BGC-Argo), and I agree with them, and these sections of the manuscript stand out as the novel scientific content. The title fails to give this message and won't promote this novel scientific content of the manuscript. I leave it to the authors consideration.*

REPLY – We agree: we changed the title highlighting the novel metrics and the usefulness of BGC-Argo float data. The title now is: “Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts”.

Reviewer #1 – Technical corrections

R1.a) (P2 L1) As a user-driven
Corrected adding “a”.

R1.b) (P3 L21) semi-labile
Corrected adding “-”.

R1.c) “In situ” appear as in-situ in various parts of the text. Ocean science journal asks for the use “in situ”.
Thanks: the revised text includes now only “in situ”.

R1.d) (P7 L16) replace relays by relies
Corrected.

R1.e) (P7 L20) remove first “run”
Removed.

R1.f) (P8 L8) Add CORR acronym
The “CORR” acronym has been added.

R1.g) (P9 L24) model simulates well
Corrected.

R1.h) (P18 L1) represent a useful
Corrected.

R1.i) (Fig10) Color code for NITR1 and NITR2 is missing
Thanks: we added the color for NITRCL1 (blue) and NITRCL2 (red). See also point R2.11.

R1.j) (Fig13) Authors used CHL for 60-100 m twice, suggesting NO3 is missing at that depth range in the figure.
Thanks: the scatter plot for nitrate at 60-100 m erroneously repeated the one for chlorophyll. Fig.13 has been corrected in the revised version (see also point R2.17).

Reviewer #2 – General comments

*R2.M1 It doesn’t give an overview of other published simulations on the Mediterranean Sea. The reader is not able to appreciate the quality of this simulation in relation to other BGC models in the literature;*

REPLY – We agree: the revised version refers to other biogeochemical model applications in Mediterranean Sea. In particular, considered the focus of the manuscript (i.e., biogeochemical operational forecast and its uncertainty estimates), we have included the Poseidon operational system (HYBRID-POM-ERSEM model, Tsiaras et al., 2017; Petihakis et al., 2018) at (RP2,L31-33):

*This is applicable to the Mediterranean Sea operational systems which include, besides CMEMS Med-MFC, also the Poseidon operational system built on the HYBRID-POM-ERSEM model coupling (Tsiaras et al., 2017; Petihakis et al., 2018).*

Besides that, there are other recent publications describing multi-annual simulations (a not exhaustive list includes Mattia et al., 2013; Macias et al., 2014; Guyennon et al., 2015; Richon et al., 2017) but we think it is out of the scope of this manuscript to list all the models and to provide an overview of their characteristics.

Nevertheless, given that the focus of our manuscript is on validation, in the revised version we have added a synthetic overview of the data availability for validating biogeochemical simulations in the Mediterranean Sea. The overview shows that the most used variable for validation is satellite-derived surface chlorophyll (Tsiaras et al., 2017; Mattia et al., 2013; Macias et al., 2014; Guyennon et al., 2015; Richon et al., 2017). Further, in situ measurements from vessels and scientific cruises are also used in Richon et al. (2017) and Guyennon et al. (2015), but allow only to validate very limited time and space portions of the simulations. Only a few basin-wide validation frameworks, especially for nutrients, are based on comparison with climatology (e.g. Tsiaras et al., 2017, used a seasonally aggregated reference for the whole Mediterranean Sea built on 1990-1999 data from SeaDataNet). Very rarely the vertical proprieties of biogeochemistry are assessed (e.g. Guyennon et al., 2015 and Teruzzi et al., 2014).

The novelty of our work lies in the introduction of the use of BGC-Argo floats data as observational counterpart for the model validation, and presents a comprehensive validation framework from basin-wide and seasonal scale to mesoscale and weekly scale (following GODAE recommendations), with an accuracy level depending on the specific variable and the availability of reference data.

The following paragraph has been added in the Introduction (RP3,L1-14):

*More in general and concerning biogeochemical applications in the Mediterranean Sea, the limited availability of observational reference data often hinders the validation assessment of model products. The most common approach is based on contrasting model outputs with satellite-derived surface chlorophyll (Tsiaras et al., 2017, for year 2000; Mattia et al., 2013; Macias et al., 2014; Guyennon et al., 2015; Richon et al., 2017, for a portion or the whole respective investigated multi-year periods). In situ measurements from vessels and scientific cruises are also used in Richon et al. (2017) and Guyennon et al. (2015), but allow only to validate very limited temporal and spatial subsets of the simulations (i.e., time series of fixed stations or single transects in a very confined time range). On the other hand, a few basin-wide validation frameworks, especially for nutrients, are based on comparison with climatology, e.g. Tsiaras et al. (2017) used a seasonally aggregated reference for the whole Mediterranean Sea built on 1990-1999 data from SeaDataNet. Generally, modelled vertical proprieties of biogeochemistry are rarely assessed (e.g. Guyennon et al., 2015 and Teruzzi et al., 2014) due to the lack of adequate reference datasets. In the recent years, the availability of biogeochemical vertical profiles in the Mediterranean Sea has significantly increased with the deployment of Biogeochemical Argo floats (hereafter BGC-Argo floats; Johnson and Claustre, 2016), whose datasets constitute an unprecedented source of reference for* *biogeochemical model skill assessment, spanning from basin-wide and seasonal scale to mesoscale and weekly scale.*

*R2.M2 It doesn't manage to rationalize the discrepancies diagnosed by the comparison with observations and to disentangle the origin of the observed biases (problems due to the physical model, the biogeochemical parameterizations or to the sensor quality control?). They are a lot of metrics that are not enough used to suggest some solutions and to improve the future systems*

REPLY – We thank the Reviewer for this comment. The submitted manuscript discussed a number of general weaknesses of the MedBFM model system (also reported in the revised version): (RP12, L3-5) refer to the high winter chlorophyll RMSD values in the nwm sub-basin, relating it with possible drawbacks in the reproduction of vertical mixing included in the sub-grid parameterization of Med-PHY component; (RP14, L5-10) refer to discrepancies in oxygen estimation, which may be related with possible inconsistency of the QC procedure of BGC-Argo oxygen data (see also R2.12); (RP19, L1-15) refer to the decreased skill at the domain boundaries and at the coastal areas. Some of these will be objects of our future development work, as already planned within Med-MFC (e.g. vertical mixing scheme in Med-PHY, oxygen formulation in Med-BIO), and also requested by the CMEMS "continuous improvement" paradigm.
In the revised version we have enriched the comments related to the results gathered from the metrics application and coherently related them with our analysis on possible causes. In particular, we have improved our discussion by reporting and rationalizing the following cases where our validation framework can detect specific discrepancies:

a) As reported at point R1.2, we highlight that our definition of the 16 sub-basins revealed to be a sensible choice, since it allowed to demonstrate the good model performance in reproducing the mean spatial gradients within the domain (Fig. 5), and the possible anomalies (as the one in nwm, see following item). We include the point related to the domain subdivision in the Discussion at (RP18,L3-5), adding a comment on the detection of the surface nitrate underestimation of Fig. 5: *[…], justifying a posteriori our sensible definition of the 16 sub-basins. In fact, the comparison of nutrients profiles of Fig. 5 highlights the satisfactory model performance in reproducing the mean spatial gradients and the possible anomalies, such as the underestimation of upper layer nitrate in nwm sub-basin.*
b) Again as in R1.2, we relate the underestimation of nitrate in the nwm sub-basin in the subsurface layer partly to possible underestimation of the terrestrial input and partly to the impact of the incoming Atlantic waters. This point is included in Section 5.1.1, at (RP10,L13-15): *Uncertainty in nwm upper layer nitrate (Fig. 5) is partly related to a possible underestimation of the Ebro/Aude/Rhone rivers input forcing and possibly to the effect of lateral circulation from Alboran Sea and Southern Western Mediterranean surface waters (see Fig. 5, panels "alb" and "swm2").*
c) Another detected discrepancy that we have included in the revised version is the presence of some overestimation for the oxygen field between model and BGC-Argo profiles below 200 m (as shown in Tab.5 and in the revised Fig. 11, with an extended value range which better illustrates the vertical gradients, see also point R2.12). Since the quality check protocol of oxygen data from BGC-Argo floats is continuously progressing, we envisage the need for a specific study regarding the oxygen validation and the analysis of oxygen variability simulated by the MedBFM model. A sentence has been added in Section 5.1 at

(RP18,L19-20): *A specific investigation focused on the oxygen validation framework and the analysis of the oxygen variability simulated by the MedBFM model is in preparation.*

d) As also reported at point R1.4, the metrics based on BGC-Argo data (e.g., DCM and MWB) are very innovative and informative. BGC-Argo floats provide simultaneous measurements of physical variables, such as profiles of PAR and temperature, that act as forcing of the biogeochemistry processes (or proxies for the forcing impacting the biogeochemistry). Thus, an integrated analysis of the model-observation uncertainty can be made comprehensively using multivariate analysis (e.g. PCA and neural network methods), with the aim to disentangle the error sources on vertical profiles. Moreover, we have to note that such analysis would need not just simple measures of distance between observations and model values (such as BIAS and RMSD), but indexes that can put in relation the shape and intensity of the biogeochemical profiles with the underlying processes (light limitation, temperature kinetic dependencies, nutrient availability through vertical fluxes, vertical dispersion through mixing). Within this framework, we think that our work (and the proposed metrics) provides a first step towards the identification and quantification of several functional indexes. Beside this consideration, another critical point is the availability of a sufficient amount of float profiles for variables like nitrate and oxygen, which would be required for a statistically significant analysis covering most of the Mediterranean regions and a longest period of the year. The proposed comment is at point R1.4, included in the Discussion at (RP18,L26-34), and here reported for sake of clarity: *In perspective, the integration of BGC-Argo within operational ocean forecasting systems in terms of data assimilation (see Cossarini et al., 2019) becomes strategical for an in-deep study of the interior of the sea and its dynamics. Moreover, considering that BGC-Argo floats also provide profiles of physical quantities (i.e., radiometric quantities – PAR – and temperature), an analysis of specific physical (e.g., MLD, euphotic zone depth) and biogeochemical (e.g., NITRCL, MWB, DCM) indexes that can reveal relationships between the shape and/or intensity of the profiles and the underlying dynamics would allow to further delve into coupled vertical physical-biogeochemical processes. In such a view, our work provides a first step to identify and quantify several functional indexes. Nonetheless, a critical point remains the availability of a sufficient amount of profiles for variables like nitrate and oxygen, which may allow for statistically significant analysis.*

e) Regarding the chlorophyll satellite-model comparison we extended the sentence in Section 5.1.1 (RP10, L3-8), adding more details on the discrepancies of the bloom model appearance in the north-western Mediterranean, with specific references, and coherently moving a sentence concerning the same issue from Section 5.1.2. Here we hypothesize a possible link with the physical dynamics, however the specific investigation is left out of the present contribution.
*Finally, modelled late winter-early spring surface chlorophyll maxima in nwm appear anticipated of 2-3 weeks w.r.t. satellite ones: this is related to a possible mismatch of the spatial patterns which characterize the temporal succession of deep convection and subsequent stratification and bloom, known to have a very high patchy (i.e., at mesoscale and sub-mesoscale) dynamics in this area (Estrada et al., 2014; Mayot et al., 2017; Severin et al., 2017). The magnitude, timing and spatial pattern of such mesoscale and sub-mesoscale structures might not be completely well resolved, thus resulting in increased discrepancies with observations.*

Reviewer #2 – Specific comments

*R2.1. P3 L26-27: "surface chlorophyll . . . of CMEMS" Which CMEMS product? Is it the same as the one used for assessment on Fig 3 and Fig 4?*

REPLY – The product is the OCEANCOLOUR_MED_CHL_L3_NRT_OBSERVATIONS_009_040 as described in section 2.2 ("Set up of the pre-operational qualification simulation for Med-BIO") and section 3 ("Reference datasets for validation").
Moreover, it is important to remind (as explained in the submitted version, and reported in the revised one at RP8,L6-8 and RP11,L29-31) that this product is used in the data assimilation scheme (at the beginning of each weekly run at day T-7) and then to assess the skill of the forecasts (as in Fig. 3 and 4 for the pre-operational run, and in Fig. 12 for the forecast run) before it is used for assimilation at the next weekly run. Following Mattern et al. (2018), the root mean square (RMS) of the differences between observation and prior model solution provides a measurement of the data assimilation performance and represents a short-term forecast skill metric since it is based on observations that are going to be assimilated in the following cycles. We think that this concept is already clearly depicted in the text. However, we are open to add details if the Reviewer suggests they are needed.

*R2.2. P4 L10: Which database do you use for river inputs? Reference?*

REPLY – Details on the boundary conditions for river nutrients inputs were reported in Section 2.2, 5th and 6th items, and reported in the revised version at (RP6,L3-11), and derive from the dataset built during the PERSEUS FP7 project. In the revised version, we include that the dataset is based on a total of 39 rivers with runoff larger than 50 $m^3/s$ (RP6,L3). To the best of our knowledge, this deliverable represents the most up-to-date information about terrestrial nutrient discharges for the Mediterranean basin.
To further improve readability, in the revised version we have added the reference to the river details described in Section 2.2 also in Section 2.1 at (RP4,L28).

*R2.3. P4 L15-16: "Additional 2D fields include the surface data for solar irradiance and wind stress": Where are these data from?*

REPLY – These data are provided via the offline coupling with MED-PHY component, which derives the atmospheric forcing from the 6-hours ECMWF operational analysis and forecast fields at 0.125° horizontal-resolution. For details please refer to CMEMS catalogue[1]. Clarifications about the source of the 2D fields have been added to the revised Section 2.1 at (RP5,L1-3):

*Additional 2D fields from MED-PHY include the surface data for solar shortwave irradiance and wind stress (derived by the ECMWF atmospheric forcing, see details later), which are used, respectively, as input for the BFM optical module and to solve the gas air-sea exchanges.*

*R2.4. P5 L11: Is the spin-up forced by a climatological year?*

---

[1] http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=MEDSEA_ANALYSIS_FORECAST_PHY_006_013

REPLY – The spin-up is based on year 2016 run without data assimilation. This procedure is coherent with that adopted for the spin-up of the pre-operational run of the physical component of the Med-MFC (see details in Clementi et al., 2018). We think it is not necessary to add further details to the revised text (RP5,L26-27), but we are open to do it in case the Reviewer will find it useful.

*R2.5. P6 L16-20: is it the same dataset as the one used in the data assimilation scheme? Can you precise the temporal frequency. Daily?*

REPLY – Yes, the dataset is the same, and it is used at daily frequency. We have added "daily" in Section 3 at (RP7-L2) to better clarify this point. As discussed at point R2.1, validation is based on satellite surface chlorophyll data that are not yet assimilated.

*R2.6. P6 L24-30: do you use the BGC-Argo dataset available on Coriolis website?*

REPLY – We download our data from LOV database due to specific quality control developed in the frame of the CMEMS Service Evolution project MASSIMILI (see details at https://www.mercator-ocean.fr/en/portfolio/massimili-2/). The use of BGC-Argo floats data from Coriolis and CMEMS INSITU-TAC is presently under consideration.

*R2.7. P11 L19: prefer 'averaged' instead of 'integrated'*

REPLY – We agree with the suggestion, we have substituted 'integrated' with 'averaged' at (RP12,L24).

*R2.8. P11 L20 and Fig9: why do the correlation values vary so abruptly and reach zero sometimes in winter?*

REPLY – We thank the Reviewer for having raised this issue. The plots of Fig. 9 (right panels in the submitted version) were erroneously computed including also dates without float data (see also R2.10). In other few cases, very low correlation values were related to some possible inconsistencies in the float measurements (see for example the isolated chlorophyll values up to 0.20 mg/m$^3$ below 200 m in April 2016 and in May 2017). Our operational check of the reliability of the input data is still progressing. Thus, in these cases, we safely decided to remove from the computation these profiles. A corrected and improved version of Fig. 9 has been included in the revised version.

*R2.9. P11 L22-23 and Fig9: This sentence is not convincing. Why these differences between modelled and observed depth of MWB? What is the difference between the mixed layer depth (MLD) and the depth of MWB? Why did you choose this index? Did you compare the mixed layer depths along the float trajectory?*

REPLY – We thank the Reviewer for this comment. We introduced the MWB in order to identify an index quantifying the thickness of the layer affected by the surface chlorophyll winter bloom. As commented in the Reply to R2.M2 (item “d”), this is a first step towards a specific metric to keep track of biogeochemical processes and their relationship with physical drivers.
As also explained in the reply to R1.1, the definition of MWB is based on the phenomenological results reported by Lavigne et al. (2015), who identified some standard shapes of chlorophyll profiles analyzing a large number of chlorophyll profiles (from fluorescence data) in the Mediterranean Sea (see their Fig. 2). Three of those profile shapes (i.e., “homogeneous”, “HSC” and “complex”) are characterized by decreasing-with-depth values and are typically observed during the winter months in different Mediterranean regions (Fig. 5 of Lavigne et al., 2015). They are not necessarily limited by the MLD, since other factors (e.g., light and nutrient conditions) play a role in the surface bloom dynamics. In addition, there are several possible definitions of MLD which can give slightly different results, so a sensitivity analysis of the biogeochemical metrics to the MLD definition would be necessary to tune the choice of the criteria based on MLD.
The specific choice of the limit of 10% was made after a sensitivity analysis varying the limit between 1 to 10% (not shown). The value of 10% gave results qualitatively consistent with those reported by Lavigne et al. (2015), whilst lower percentage values gave more unclear patterns because the depth increases substantially and the thickness of model layers has also an impact, due to the vertical resolution coarsening (i.e. thickness of model layers is 5 m at 70 m depth, 7 m at 100 m and 10 m at 140 m). Further, the MWB metric fails to capture the “modified DCM” shape profile (as defined in Lavigne et al., 2015) which occurred for float 6901653 during winter 2016 (Fig. 9). Hence, given the constraint in the definition and the vertical discretization of the model, the application of this index to the floats data may originate some under- or overestimations and uncertainty of a few decameters (see Tab. 3). Despite these limitations, we think that the MWB represents a feasible and informative metric to be coupled with the DCM metrics.
In the revised manuscript, we have better explained the scientific rationale of the MWB metrics in Section 4.1 (RP8,L27-32 and RP9,L1-2), here also reported for sake of clarity:

*The definitions of DCM and MWB metrics are consistent with the outcomes of Lavigne et al. (2015), who identified some standard shapes for chlorophyll vertical profiles and their temporal distribution from the analysis of a large dataset of fluorescence data in the Mediterranean Sea (see their Figs. 2 and 5). In particular, the summer period defined to estimate the DCM index is based on the consideration that the DCM profile shape is typically observed from April to October. Otherwise, the choice to limit the estimate of the MWB index from January to March is motivated by the fact that steady depth-decreasing profiles typically occur during that period in different Mediterranean regions. Further, the choice of the 10% criterion for the MWB index was set after a sensitivity analysis varying the threshold between 1 to 10% (not shown), with the 10% value giving results qualitatively consistent with those reported by Lavigne et al. (2015).*

Moreover, we have modified the original sentence in Section 5.1.2 according to this comment (RP13,L1-6):

*Considering the constraint in the definition of the MWB depth and the vertical discretization of the model, the application of such index to floats data may originate some inconsistency (as shown for winter 2016 in Fig. 9), and under- or overestimations and uncertainty of a few decameters (see Tab. 3). Despite these limitations, we consider the MWB as a feasible and informative metric alongside the DCM metrics to characterize the seasonal chlorophyll profile evolution.*

*R2.10. Fig9: in summer 2017, there is no diagnosed DCM for model. Why?*

REPLY – We thank the Reviewer: there was an error in the plot (see also point R2.8). The revised version of Fig. 9 reports a wider y-axis in the bottom panel, and we observe that the DCM in 2017 appears shallower than in 2016 (please also consider the float trajectory – in the top panel of Fig. 9 – that covers the south-western Mediterranean from Balearic Islands to Alboran Sea, which is known to have a shallower DCM, see Lazzari et al., 2012).

*R2.11. Fig 10: skill index 4th row: the legend is missing for NITRCL1/NITRCL2. Why is the blue index missing for model in the beginning of 2017?*

REPLY - We thank the Reviewer: we have corrected the caption (NITRCL1 is in blue and NITRCL2 is in red). We have also extended the maximum depth to 300 m (new version of Fig. 10), which allows to recognize the agreement between NITRCL1 values computed on model and BGC-Argo data at the beginning of 2017. It can be observed that in the period April-July 2017 one of the two nitracline indexes (i.e., NITRCL2) computed on model output is much shallower than the one estimated from the BGC-Argo float data. Thus, the two indexes were useful to show different aspects of the nitrate profile evolution, justifying their use to provide indications aimed to monitor the model error behaviour.
A comment has been added in the revised version in Section 5.1 at (RP13,L16-19):

*The model NITRCL1 and 2 perform generally good, however, it can be observed that in the period April-July 2017 the NITRCL2 appears much shallower than what estimated by the float data. The two indexes show different aspects of the nitrate profile evolution, justifying their use to provide indications aimed to monitor the model error behaviour.*

*R2.12. Fig10-11: Authors don't show depths below 300m. For oxygen, the colorbar is saturated below 200 umol/L. It doesn't allow to study the nitrate gradient and the OMZs.*

REPLY – The model comparison for depths below 300 m is shown in Fig. 5 for nutrients and oxygen for 6 selected sub-basins. Given the slow dynamics of the deep layers, the comparison based on climatology (GODAE Class 1 metrics with NODC-OGS dataset) demonstrates the capability of the model to reproduce the nutrient values along the Mediterranean Sea gradient and the relative minimum of the oxygen concentration below 400 m.
The depth of the relative minimum of oxygen displayed in Fig. 5 is consistent with published information: according to Tanhua et al. (2013; see their Fig. 6), the oxygen minimum layer (OML) core in Mediterranean is located at 500 - 700 m in the eastern basin, "well below the layer of maximum S [salinity] occupied by the LIW" (see sub-basins ion2, ion3 and lev4 in Fig. 5), whilst in the western basin it is shallower at around 400 m depth (see sub-basin nwm in Fig. 5; other sub-basins can be found in Cossarini et al., 2018). A sentence about such agreement with the historical data has been added in Section 5.1 at (RP10,L27-30):

*We can observe that the depth of the relative minimum of oxygen displayed in Fig. 5 is consistent with the cruise data shown by Tanhua et al. (2013): the oxygen minimum layer core in the eastern basin is located below 500 m (sub-basins ion2, ion3 and lev4 in Fig. 5), whilst is at around 400 m* *depth in the western basin it (see sub-basin nwm in Fig. 5; for the sub-basins please refer to Cossarini et al., 2018).*

On the other hand, BGC-Argo floats allow to observe relatively fast dynamics (i.e., at weekly time scale according to their current 5-day sampling frequency) which are particularly relevant for the biogeochemical processes in the upper layer. Thus, we used BGC-Argo floats data to verify the model capability to reproduce the mesoscale and vertical dynamics in the upper layer at the weekly time scale. This represents a very stringent comparison, which allowed to demonstrate the timing and intensity of the evolution of signals that track biogeochemical processes in the upper layer. Therefore, we think that it would be not enough illustrative to compare and show deep profiles for model and float, also considering the ongoing and progressing advancements of the quality control process for BGC-Argo floats data. In fact, only very recently a new product quality control procedure (following Bittig et al., 2018 and Thierry et al., 2018) has been started to be implemented for oxygen to correct biases on sensor: to the best of our knowledge, it is not yet available for all floats in the Mediterranean Sea. Thus, we prefer to keep the maximum depth of Figs. 10 and 11 at 300 m to better exploit the model-float visual comparison in the upper layer (roughly identified with the euphotic layer, generally not deeper than 200 m), where most of the biogeochemical processes occur.
We have revised the colorbar of Fig. 11 starting from oxygen value at 180 mmol/m$^3$: the new Figure 11 displays some discrepancies (i.e., a bias of about 10-20 mmol/m$^3$) that were already commented along with Table 5. However, given the aforementioned ongoing quality improvements in the BGC-Argo float oxygen data, a quantitative calculation of the discrepancy between model and floats (as done for the chlorophyll and nitrate metrics) has not been defined yet, and we proposed only qualitative considerations, as reported at (RP14, L5-10).
In the revised manuscript, we have better clarified the limitations of the BGC-Argo data use for oxygen, adding a sentence in Section 5.1 (RP14,L10-12):

*Only very recently a new product quality system (following Bittig et al., 2018 and Thierry et al., 2018) has started to be implemented for oxygen data to correct biases on sensor: to the best of our knowledge, it is not yet available for all floats in the Mediterranean Sea.*

Finally, we would like to point out that with improving oxygen data quality, the use of BGC-Argo floats to investigate quantitatively the simulated oxygen will become very promising for evaluating surface and deeper layer dynamics. A comment on the improvement of the oxygen validation framework has been added in the discussion regarding future developments at (RP18,L19-20), as already proposed at item “c” of R2.M2:

*A specific investigation focused on the oxygen validation framework and the analysis of the oxygen variability simulated by the MedBFM model is in preparation.*

*R2.13. P12 L21: I would say ‘solubility’ instead of ‘saturation’*

REPLY – Thanks, the term “solubility” has been used in the revised version.

*R2.14. P12 L22: What are the consumption terms?*

REPLY – Comsumption terms are defined as respiration terms by bacteria and plankton community (4 phytoplankton and 4 zooplankton groups). We have added the definition in the revised version (RP13,L33-34):

*[…] presence of consumption terms (defined as respiration terms by bacteria and plankton community: 4 phytoplankton and 4 zooplankton groups).*

*R2.15. Fig 12: needs to be improved (quality). Do you compare to the daily NRT L3 chlorophyll product? The figure seems to be at a coarser temporal resolution. Moreover, the yaxis labels don't appear correctly (they are truncated). I don't understand why the red curve is constant during the summer.*

REPLY – We agree about the poor quality of Fig.12: an updated version (using bar plot and extending the time range until January 2019) is present in the revised document. The comparison is made, for each week, between the first 3 days of forecast (black for T1, blue for T2 and green for T3) and the corresponding daily NRT L3 satellite product (as written at RP14,L24). The red line represents the seasonal benchmark defined as the averages of the RMSD shown in Fig. 8. Thus, the 2 seasonal benchmarks give information on the average quality of forecast for the 2 periods.

*R2.16. P13 L15: 0.041 instead of 0.41 mg/m3?*

REPLY – Thanks: yes, it is 0.041 mg/m$^3$. It has been corrected in the revised version (RP14,L29).

*R2.17. P13 L18-20 and Fig 13: Authors don't comment much the Fig13. For example, the nitrate dots are very scattered (between 0 and 100m depth), do you think that it is due to nitrate sensor anomalous values or is it a problem in the physical or biological models? Oxygen panels display a bias (model overestimates float data): do you think it is a bias in the model or in the data?*

REPLY – We acknowledge the Reviewer for this point: the comment to Fig. 13 has been further enriched in the revised version (RP15,L3-12). As we pointed out in the Discussion of the submitted manuscript (reported in the revised one P18,L16-17), and also thoroughly commented at R2.12, the systematic bias for oxygen at depth can be either due to both model and data uncertainty. Only very recently, new product quality procedure (following Bittig et al., 2018 and Thierry et al., 2018) has started to be implemented, but up to now is not yet available for all floats in the Mediterranean.
On the other hand, concentrations of oxygen lower than 180 mmol/m$^3$ in subsurface layer (100-150 m) appear quite suspicious for the Mediterranean Sea (see Manca et al., 2004, and also Tanhua et al., 2013).
An in-depth analysis of the oxygen vertical profile dynamics is therefore preliminary. The main message we would like to convey here is the methodological approach that can be used to keep monitored the NRT forecast w.r.t. to a benchmark provided by a past simulation. We have added in Section 5.2, at (RP15,L3-4) the following sentence, which introduces Fig. 13:

*To provide a monitoring of the quality of the NRT forecast with respect to a seasonal reference defined by the pre-operational qualification run, […]*

And, at (RP15,L10-12):

*We can observe that floats oxygen concentrations in the subsurface layer (100-150 m) are lower than 180 mmol/m$^3$, which appears quite anomalous for the Mediterranean Sea (see Manca et al., 2004, and also Tanhua et al., 2013), thus conveying a suspect instrumental bias of the oxygen sensor, as already discussed in Section 5.1.2 for Fig. 11.*

Regarding nitrate, the scatter plot at 60-100 m erroneously repeated the one for chlorophyll: in the corrected version we observe that the operational forecasts are generally in line with the seasonal benchmark (i.e. most of the numbers are within the orange points cloud).

*R2.18. Table 6: I am wondering why the chlorophyll RMSD for the pre-operational is twice as large as the one for the T0 forecast. Do you have an explanation for the nitrate RMSD decrease from T0 to T3?*

REPLY – We thank the Reviewer for this comment, since it helped to better discuss the results of Tab. 6. The limited observations available during the operational period (less than 5 per week) may hinder the statistical significance. In particular for chlorophyll, the qualification period spans 2 years, while the operational data considered for Tab. 6 ranges from April to October, when the model errors are much lower (see Fig. 4).
For what concern nitrate, the decrease of its RMSD is related to the limited amount of data available (BGC-Argo floats data may exhibit wide oscillations over subsequent profiles, as shown in Fig. 10). The important aspect here is that the quality of biogeochemical variables in the first 4 forecast days ranges within ±25%, which allows to conclude that the quality of biogeochemical forecast does not significantly degrade and remains satisfactory (i.e., in line with the benchmark) during the first week.
Following such considerations: the comment to Tab. 6 has been mostly revised (RP15,L13-18):

*The RMSDs of the four forecast days (Tab. 6) remain within a range of ±25%, generally showing that the quality of biogeochemical forecast does not significantly degrade during the first week. More precisely, chlorophyll and oxygen RMSD of T3 and T4 are slightly larger than T1, while nitrate RMSD of the last forecast days is lower than T1. However, considering the very low number of available data (few tens in the 5 months considered) and the fact that BGC-Argo floats data may exhibit wide oscillations over subsequent profiles (as shown in Fig. 10), the differences of skill performance statistics from one day of forecast to another might be considered cautionary.*

Finally, we have also stressed in the Conclusions (RP20,L12-14) that robust statistics require much longer time series of data and a larger number of BGC-Argo floats, which is becoming an urgent request for the observing systems to be used in operational biogeochemical oceanography (for both validation and assimilation purposes):

*Robust statistics require much longer time series of data and a larger number of BGC-Argo floats, which is becoming an urgent request for the observing systems to be used in operational biogeochemical oceanography (for both validation and assimilation purposes).*

Reviewer #2 – Technical comments

R2.a) P4 L16-17: "are used respectively . . .module". I think you should switch both: "as input for the BFM optical module and to solve the gas air-sea exchanges"
We agree, text has been corrected.

R2.b) P4 L21: temperature AND salinity
We agree, we have also included a comma before "and along-track".

R2.c) P6 L14: the same AS
Corrected.

R2.d) P7 L2: SchmeCHtig
Corrected.

R2.e) P7 L16: relies instead of relays
Corrected.

R2.f) P9 L6-7: surface bloomS in nwm appear (remove the s)
The sentence has been modified, according to R2.M2.

R2.g) P9 L9-10: NODC-OGS instead of OGS-NODC for more consistency with the beginning of the manuscript
Corrected.

R2.h) P10 L7: "a large seasonal cycles": remove the "s" in cycles
Corrected.

R2.i) P11 L10 "river plume" : remove the "s"
Corrected.

R2.j) P18 L1: represent a (remove the 'n')
Corrected.

R2.k) Fig 14: in caption, 'increase' instead of 'increment'?
Yes, we modified with "increase".

## References

Bittig, H. C., Körtzinger, A., Neill, C., van Ooijen, E., Plant, J. N., Hahn, J., Johnson, K. S., Yang, B., Emerson, S. R. (2018). Oxygen Optode Sensors: Principle, Characterization, Calibration, and Application in the Ocean. Frontiers in Marine Science 4: | DOI: 10.3389/fmars.2017.00429

Clementi, E., Grandi, A., Di Pietro, P., Pistoia, J., Delrosso, D., Mattia, G. (2018). 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

Cossarini, G., Salon, S., Bolzon, G., Teruzzi, A., Lazzari, P., and Feudale, L. (2018). 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.

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

Estrada, M., Latasa, M., Emelianov, M., Gutiérrez-Rodríguez, A., Fernández-Castro, B., Isern-Fontanet, J., Mouriño-Carballido, B., Salat, J. and Vidal, M.: Seasonal and mesoscale variability of primary production in the deep winter-mixing region of the NW Mediterranean. Deep Sea Research Part I: Oceanographic Research Papers, 94, pp.45-61, 2014.

Guyennon, A., Baklouti, M., Diaz, F., Palmieri, J., Beuvier, J., Lebaupin-Brossier, C., Arsouze, T., Béranger, K., Dutay, J.-C., and Moutin, T. (2015). New insights into the organic carbon export in the Mediterranean Sea from 3-D modeling, Biogeosciences, 12, 7025-7046, https://doi.org/10.5194/bg-12-7025-2015.

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.

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, 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.

Macias, D., Stips, A., and Garcia-Gorriz, E. (2014). The relevance of deep chlorophyll maximum in the open Mediterranean Sea evaluated through 3d hydrodynamic-biogeochemical coupled simulations. Ecological Modelling, 281:26-37.

Manca, B., Burca, M., Giorgetti, A., Coatanoan, C., Garcia, M.J., Iona, A.: Physical and biochemical averaged vertical profiles in the Mediterranean regions: an important tool to trace the climatology of water masses and to validate incoming data from operational oceanography, J. Mar. Syst. 48, 83–116, https://doi.org/10.1016/j.jmarsys.2003.11.025, 2004.

Mattern, J.P., Edwards, C.A., Moore, A.M. (2018). Improving variational data assimilation through background and observation error adjustments. Mon. Weather Rev. 146 (2), 485–501

Mattia, G., Zavatarelli, M., Vichi, M., and Oddo, P. (2013). The eastern Mediterranean Sea biogeochemical dynamics in the 1990s: A numerical study. Journal of Geophysical Research: Oceans, 118(4):2231-2248.

Mayot, N., D'Ortenzio, F., Taillandier, V., Prieur, L., Pasqueron de De Fommervault, O., Claustre, H., Bosse, A., Testor, P. and Conan, P.: Physical and biogeochemical controls of the phytoplankton blooms in North Western Mediterranean Sea: A multiplatform approach over a complete annual cycle (2012–2013 DEWEX experiment). Journal of Geophysical Research: Oceans, 122(12), pp.9999-10019, 2017.

Mignot, A., D'Ortenzio, F., Taillandier, V., Cossarini, G., & Salon, S. (2019). Quantifying observational errors in Biogeochemical-Argo oxygen, nitrate, and chlorophyll a concentrations. Geophysical Research Letters, 46. https://doi.org/10.1029/2018GL080541

Petihakis, G., Perivoliotis, L., Korres, G., Ballas, D., Frangoulis, C., Pagonis, P., Ntoumas, M., Pettas, M., Chalkiopoulos, A., Sotiropoulou, M., Bekiari, M., Kalampokis, A., Ravdas, M., Bourma, E., Christodoulaki, S., Zacharioudaki, A., Kassis, D., Potiris, E., Triantafyllou, G., Tsiaras, K., Krasakopoulou, E., Velanas, S., and Zisis, N.: An integrated open-coastal biogeochemistry, ecosystem and biodiversity observatory of the eastern Mediterranean – the Cretan Sea component of the POSEIDON system, Ocean Sci., 14, 1223-1245. https://doi.org/10.5194/os-14-1223-2018, 2018.

Richon, C., Dutay, J.-C., Dulac, F., Wang, R., Balkanski, Y., Nabat, P., Aumont, O., Desboeufs, K., Laurent, B., Guieu, C., et al. (2017). Modeling the impacts of atmospheric deposition of nitrogen and desert dust-derived phosphorus on nutrients and biological budgets of the mediterranean sea. Prog. Ocean. 163, 21 - 39. https://doi.org/10.1016/j.pocean.2017.04.009

Severin, T., Kessouri, F., Rembauville, M., Sánchez-Pérez, E.D., Oriol, L., Caparros, J., Pujo-Pay, M., Ghiglione, J.F., D'Ortenzio, F., Taillandier, V. and Mayot, N.: Open-ocean convection process: A driver of the winter nutrient supply and the spring phytoplankton distribution in the Northwestern Mediterranean Sea. Journal of Geophysical Research: Oceans, 122(6), pp.4587-4601, 2017.

She, J., Allen, I., Buch, E., Crise, A., Johannessen, J. A., Le Traon, P.-Y., Lips, U., Nolan, G., and other 5 co-authors: 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.

Tanhua, T., Hainbucher, D., Schroeder, K., Cardin, V., Álvarez, M., and Civitarese, G. (2013). The Mediterranean Sea system: a review and an introduction to the special issue, Ocean Sci., 9, 789-803, https://doi.org/10.5194/os-9-789-2013.

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, 2018.

Thierry V., Bittig H., The Argo-Bgc Team (2018). Argo quality control manual for dissolved oxygen concentration. https://doi.org/10.13155/46542

Tsiaras, K. P., Hoteit, I., Kalaroni, S., Petihakis, G., and Triantafyllou, G. (2017). A hybrid ensemble-OI Kalman filter for efficient data assimilation into a 3-D biogeochemical model of the Mediterranean. Ocean Dynamics, 67(6):673-690.

# Novel metrics based on Biogeochemical Argo data to improve the model uncertainty evaluation of the CMEMS Mediterranean marine ecosystem forecasts

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

[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 and literature. We show that the quality of the MedBFM system has been improved in the previous 10 years. We demonstrate that a set of metrics based on the GODAE paradigm can be efficiently applied to validate an operational model system for biogeochemical and ecosystem forecasts. The accuracy of the CMEMS biogeochemical products for the 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, the latter being an important prerequisite to build robust statistics. In particular, the use of the Biogeochemical Argo floats data proved to significantly enhance the validation framework of operational biogeochemical models. New skill metrics, aimed to assess key biogeochemical processes and dynamics (e.g. deep chlorophyll maximum depth, nitracline depth), can be easily implemented to routinely monitor the quality of the products and highlight possible anomalies through the comparison of NRT forecast with pre-operationally defined seasonal benchmark. Feedbacks to the observing autonomous systems in terms of QC and deployment strategy are also discussed.

## 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 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 strongly related to the development of data assimilation capacity, while their quality assessment is constrained by the availability of reference data, both remote and in situ (Gehlen et al., 2015), and possibly 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 a 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. The CMEMS products are delivered to users through a service portfolio, where the information is organized by 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 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), including also data assimilation (Teruzzi et al., 2014; Storto et al., 2015).

More specifically, the last major upgrade of Med-BIO focused on the increase of horizontal resolution from 1/16 to 1/24 degree. The upgrade also 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 of surface chlorophyll concentration to coastal areas (Teruzzi et al., 2018) and reducing the time-to-solution through a parallelization of the cost function solver (Teruzzi et al., 2019).

Both historical and near-real time data of observation systems are strategical to evaluate the quality of operational oceanography products (She et al., 2016). However, 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 operational observation systems (Bell et al., 2015; Hernandez et al., 2018). 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., 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). This is applicable to the Mediterranean Sea operational systems which include, besides CMEMS Med-MFC, also the Poseidon operational system built on the HYBRID-POM-ERSEM model coupling (Tsiaras et al., 2017; Petihakis et al., 2018).

More in general and concerning biogeochemical applications in the Mediterranean Sea, the limited availability of observational reference data often hinders the validation assessment of model products. The most common approach is based on contrasting model outputs with satellite-derived surface chlorophyll (Tsiaras et al., 2017, for year 2000; Mattia et al., 2013, Macias et al., 2014, Guyennon et al., 2015, Richon et al., 2017, for a portion or the whole investigated multi-year periods). In situ measurements from vessels and scientific cruises are also used in Richon et al. (2017) and Guyennon et al. (2015), but allow only to validate limited temporal and spatial subsets of the simulations (i.e., time series of fixed stations or single transects in a very confined time range). On the other hand, a few basin-wide validation frameworks, especially for nutrients, are based on comparison with climatology, e.g. Tsiaras et al. (2017) used a seasonally aggregated reference for the whole Mediterranean Sea built on 1990-1999 data from SeaDataNet. Generally, modelled vertical proprieties of biogeochemistry are rarely assessed (e.g. Guyennon et al., 2015 and Teruzzi et al., 2014) due to the lack of adequate reference datasets. In the recent years, the availability of biogeochemical vertical profiles in the Mediterranean Sea has significantly increased with the deployment of Biogeochemical Argo floats (hereafter BGC-Argo floats; Johnson and Claustre, 2016), whose datasets constitute an unprecedented source of reference for biogeochemical model skill assessment, spanning from basin-wide and seasonal scale to mesoscale and weekly scale.

In the present paper, we focus on the CMEMS Mediterranean biogeochemical analysis and forecast system products (delivered from April 2018) and we introduce novel skill metrics based on the comparison between model products and the BGC-Argo floats data. According to the definition adopted within Copernicus community (Hernandez et al., 2018), our model validation follows two main tasks:

1. The pre-operational qualification, that 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 of forecast products, that is performed operationally based on the available NRT observations and provides an evaluation of the skill performance of the analysis and forecast 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 including 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 and macro-nutrients (nitrogen, phosphorus, and silicon) 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, semi-labile 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 CO2 (pCO2) and air-sea CO2 flux.

3DVarBio is the variational data assimilation scheme for the update 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 ($V_V$), the horizontal covariance ($V_H$) and the covariance among biogeochemical variables ($V_b$). $V_V$ 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., 2018), through the upgrade of the non-homogeneous $V_V$ and the non-uniform and direction-dependent $V_H$ specifically focussed for the case 2 waters. Further, the time-to-solution of 3DVarBio has been significantly reduced 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 (Teruzzi et al., 2019).

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, see Sect. 2.2 for details) 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 from

Med-PHY include the surface data for solar shortwave irradiance and wind stress (derived by the ECMWF atmospheric forcing, see details below), which are used, respectively, as input for the BFM optical module and to solve the gas air-sea exchanges.

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 and 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. (2018).

The analysis and forecast 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).

### 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

---

[1] MEDSEA_ANALYSIS_FORECAST_PHY_006_013
[2] MEDSEA_ANALYSIS_FORECAST_BIO_006_014
[3] OCEANCOLOUR_MED_CHL_L3_NRT_OBSERVATIONS_009_040

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 (with runoff larger than 50 $m^3/s$) 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 constant 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. A similar approach holds for the Dardanelles, considered as a river input: the total inflow was derived considering typical water mass concentration of ALK and DIC for Marmara Sea (Copin-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 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 pCO2 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 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 as the pre-operational qualification run, which provides the initial conditions of the operational system at 1 January 2018.

---

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

## 3 Reference datasets for validation

Chlorophyll data are derived by the multi-sensor (MODIS-AQUA and NPP-VIIRS) CMEMS daily product of ocean colour observations for Mediterranean Sea (see Sect. 2.2; Volpe et al., 2007, 2012, 2017) at 1 km spatial resolution. The chlorophyll field combines the estimates of two algorithms for open ocean (case 1) and coastal (case 2) water types. 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 BGC-Argo floats dataset whose 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 and Thierry (2016).

BGC-Argo chlorophyll (Chl) adjusted data are derived from real time (RT) data with a series of corrections: the quenching 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 a linear regression constrained to pass through the origin between percent oxygen solubility values derived from the float profiles of O2 and climatological values from the World Ocean Atlas Climatology (Takeshita, 2013; Schmechtig and Thierry, 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 follows 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 relies 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-year pre-operational run described in Section 2.2. In particular, the validation consists of "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 "Class 4" metrics, which quantify the differences between model and observations at their location and time ("match-ups").

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 root mean square 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 (CORR) between each couple of chlorophyll (oxygen and nitrate) vertical 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).

The definitions of DCM and MWB metrics are consistent with the outcomes of Lavigne et al. (2015), who identified some standard shapes for chlorophyll vertical profiles and their temporal distribution from the analysis of a large dataset of fluorescence data in the Mediterranean Sea (see their Figs. 2 and 5). In particular, the summer period defined to estimate the DCM index is based on the consideration that the DCM profile shape is typically observed from April to October. Otherwise, the choice to limit the estimate of the MWB index from January to March is motivated by the fact that steady depth-decreasing profiles typically occur during that period in different Mediterranean regions. Further, the choice of the 10% criterion for the

MWB index was set after a sensitivity analysis varying the threshold between 1 to 10% (not shown), with the 10% value giving results qualitatively consistent with those reported by Lavigne et al. (2015).

The rationale behind the nitracline depth metrics is defining an index useful to track the time evolution of the nitrate profile. Being aware that the choice of a specific value of nitrate concentration may be controversial, we propose two different indexes: the first is based on the depth of the 2 mmol/m$^3$ concentration isopleth (NITRCL1), the second is related to the depth of the maximum nitrate vertical gradient (NITRCL2). According to Manca et al. (2004), the values of nitrate concentration at depth higher than 400 m are around 4-5 mmol/m$^3$ in the eastern basin and 6-7 mmol/m$^3$ in the western, therefore the 2 mmol/m$^3$ isopleth can be considered a consistent threshold to detect the rapid change between the very low concentration typically measured at the surface and the high concentration at depth in all areas of the Mediterranean Sea.

### 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 and third day of forecast (i.e. forecast lead time of T1, T2 and T3) is compared with the corresponding 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 T1 to T4) 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 assessment, which acts 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 Gibraltar Strait. Finally, modelled late winter-early spring surface chlorophyll maxima in nwm appear anticipated of 2-3 weeks w.r.t. satellite ones: this is related to a possible mismatch of the spatial patterns which characterize the temporal succession of deep convection and subsequent stratification and bloom, known to have a very high patchy (i.e., at mesoscale and sub-mesoscale) dynamics in this area (Estrada et al., 2014; Mayot et al., 2017; Severin et al., 2017). The magnitude, timing and spatial pattern of such mesoscale and sub-mesoscale structures might not be completely well resolved, thus resulting in increased discrepancies with observations.

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 NODC -OGS climatological profiles (Fig. 5), and the correlation values are generally larger than 0.9 (Fig. 7), which corroborates the 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. Uncertainty in nwm upper layer nitrate (Fig. 5) is partly related to a possible underestimation of the Ebro/Aude/Rhone rivers input and possibly to the effect of lateral circulation from Alboran Sea and Southern Western Mediterranean surface waters (see Fig. 5, panels “alb” and “swm2”).

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 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.

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 RMSDs normalized by the standard deviation range between 0.60 and 1.40 at the different layers, but considering the surface temporal seasonal cycle the normalized RMSD is 0.15. We can observe that the depth of the relative minimum of oxygen displayed in Fig. 5 is consistent with the cruise data shown by Tanhua et al. (2013): the oxygen minimum layer core in the eastern basin is located below 500 m (sub-basins ion2, ion3 and lev4 in Fig. 5), whilst in the western basin (see sub-basin nwm in Fig. 5; for the other sub-basins please refer to Cossarini et al., 2018).

Figure 6 shows that the model simulates well the vertical structure of DIC and ALK, mostly within the range of variability of 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). 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 (namely: 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.

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 cycle 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).

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 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 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 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/m$^3$ 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, 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, as discussed in Section 5.1.1. 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 -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., 2018) 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 model capability to simulate bloom events triggered by river plume events (Teruzzi et al., 2018).

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 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 new quantitative metrics (defined in Sect. 4.1) computed on the vertical profiles comparison are shown in the lower panels 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 averaged 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, and the model MWB depth well performs in 2017, while it appears shallower in 2016.

Averaging the time series of RMSD and BIAS of the new metrics for the aggregated sub-basins (Tab. 3) highlights that the MedBFM model has a very high skill in reproducing the vertical dynamics of the phytoplankton chlorophyll in the 0-200 m layer, considering both 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 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 depth of the MWB is not computable and reliable for some of the sub-basins. However, for those having more than 10 float profiles per month, it has an absolute BIAS ranging from 30 to 40 m and a RMSD ranging from 40 to 50 m. Considering the constraint in the definition of the MWB depth and the vertical discretization of the model, the application of such index to floats data may indeed originate some inconsistency (as shown for winter 2016 in Fig. 9), and under- or overestimations and uncertainty of a few decametres (see Tab. 3). Despite these limitations, we consider the MWB as a feasible and informative metric alongside the DCM metrics to characterize the seasonal chlorophyll profile evolution.

The averaged vertical values show that the model generally underestimates the content of chlorophyll with respect the BGC-Argo floats measurements, which 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. (2019), 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 comparison of model nitrate with the BGC-Argo float measurements 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 model NITRCL1 and 2 perform generally good, however, it can be observed that in the period April-July 2017 the NITRCL2 appears much shallower than what estimated by the float data. The two indexes show different aspects of the nitrate profile evolution, justifying their use to provide indications aimed to monitor the model error behaviour.

Tab. 4 shows the nitrate metrics of the 8 floats, averaged over the aggregated sub-basins: 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 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, 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 solubility and presence of consumption terms (defined as respiration terms by bacteria and plankton community: 4 phytoplankton and 4 zooplankton groups). 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, 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 solubility 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°C and 0.01, respectively (Clementi et al., 2018), and under the hypothesis of oxygen solubility at surface, the BIAS for the modelled 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. Only very recently a new product quality system (following Bittig et al., 2018 and Thierry et al., 2018) has started to be implemented for oxygen data to correct biases on sensor: to the best of our knowledge, it is not yet available for all floats in the Mediterranean Sea. 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 are performed using GODAE Class 4 metrics with available satellite data for the first three days of forecast (T1-T3 lead time) and BGC-Argo floats observations for the first four days of forecast (T1-T4 lead time) using the same metrics computed for the pre-operational run, that provides the benchmarks of the accuracy level. Online resources for such metrics (RMSD and BIAS between model and observations averaged over the sub-basins) are updated quarterly on the official CMEMS validation webpage[5] and weekly on the regional Mediterranean validation website managed by OGS[6].

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 and third day of forecast. Indeed, the average of the RMSD over the 16 sub-basins is 0.018, 0.034 and 0.041 mg/m$^3$ for the

---

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

[6] Available at http://medeaf.inogs.it/nrt-validation/

first, second and third day of forecast, respectively. The high variability of the RMSD statistics from one day to another is basically related to the daily varying number of available pixels, due to the cloud cover and its spatial distribution.

To provide a monitoring of the quality of the NRT forecast with respect to a seasonal reference defined by the pre-operational qualification run, 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 by the results from the pre-operational run. In general, the forecast data are within the variability of the seasonal benchmark (in this case, the period from May to August). 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. We can observe that floats oxygen concentrations in the subsurface layer (100-150 m) are lower than 180 mmol/m$^3$, which appears quite anomalous for the Mediterranean Sea (see Manca et al., 2004, and also Tanhua et al., 2013), thus conveying a suspect instrumental bias of the oxygen sensor, as already discussed in Section 5.1.2 for Fig. 11.

The RMSDs of the four forecast days (Tab. 6) remain within a range of ±25%, generally showing that the quality of biogeochemical forecast does not significantly degrade during the first week. More precisely, chlorophyll and oxygen RMSD of T3 and T4 are slightly larger than T1, while nitrate RMSD of the last forecast days is lower than T1. However, considering the very low number of available data (few tens in the 5 months considered) and the fact that BGC-Argo floats data may exhibit wide oscillations over subsequent profiles (as shown in Fig. 10), the differences of skill performance statistics from one day of forecast to another might be considered cautionary.

## 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., 2018, 2019) and the non-linear 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-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, the increased horizontal resolution, and 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 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). Further improvements of the Med-BIO biogeochemical model system in terms 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 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 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 solubility.

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 16 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 dataset 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 related 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 a seasonal 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.

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 (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.

The present validation framework uses an *a priori* subdivision which considers the biogeographic approach of D'Ortenzio and Ribera d'Alcalà (2009), and the subsequent refinement proposed by Lazzari et al. (2012) which showed different characterizations according to the 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

---

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

provide model uncertainty 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-basins allows to synthetize the heterogeneity of the Mediterranean Sea, justifying *a posteriori* our sensible definition of the 16 sub-basins. In fact, the comparison of nutrients profiles of Fig. 5 highlights the satisfactory model performance in reproducing the mean spatial gradients and the possible anomalies, such as the underestimation of upper layer nitrate in nwm sub-basin. Moreover, the use of simple indexes, such as means and standard deviation, and of functional spatio-temporal subdivisions increase the readability of the uncertainty communication, which responds also to the request for a user oriented evolution of the validation framework in operational systems (Hernandez et al., 2018).

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., 2018).

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 seasonal benchmarking might be beneficial for an additional QC procedure for 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 data in the Mediterranean Sea. A specific investigation focused on the oxygen validation framework and the analysis of the oxygen variability simulated by the MedBFM model is in preparation.

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). In perspective, the integration of BGC-Argo within operational ocean forecasting systems in terms of data assimilation (see Cossarini et al., 2019) becomes strategical for an in-deep study of the interior of the sea and its dynamics. Moreover, considering that BGC-Argo floats also provide profiles of physical quantities (i.e., radiometric quantities – PAR – and temperature), an analysis of specific physical (e.g., MLD, euphotic zone depth) and biogeochemical (e.g., NITRCL, MWB, DCM) indexes that can reveal relationships between the shape and/or intensity of the profiles and the underlying dynamics would allow to further delve into coupled vertical physical-biogeochemical processes. In such a view, our work provides a first step to identify and quantify several functional biogeochemical indexes. Nonetheless, a critical point remains the availability of a sufficient amount of profiles for variables like nitrate and oxygen, which may allow for statistically significant analysis.

Concerning the weakness of MedBFM, we may point out 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 Raghukumar 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., 2018). 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 datasets (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 a 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 vertical profiles of 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 slight decrease of forecast skill performance after 1 or 2 days for surface chlorophyll and a not unique identified pattern when BGC-Argo data are used. Nevertheless, the forecast skill performance remains at the same level as the benchmark within the first week of forecast.

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. Robust statistics require much longer time series of data and a larger number of BGC-Argo floats, which is becoming an urgent request for the observing systems to be used in operational biogeochemical oceanography (for both validation and assimilation purposes).

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. The authors thank Dr. V. Di Biagio for fruitful discussion concerning the review of biogeochemical model applications in Mediterranean Sea.

## References

Artuso, F., Chamard, P., Piacentino, S., Sferlazzo, D.M., De Silvestri, L., di Sarra, A., Meloni, D., Monteleone, F.: Influence of transport and trends in atmospheric CO2 at Lampedusa. Atmos. Environ. 43 (19), 3044–3051, 2009.

Álvarez, M., Sanleón-Bartolomé, H., Tanhua, T., Mintrop, L., Luchetta, A., Cantoni, C., Schroeder, K., and Civitarese, G.: The CO2 system in the Mediterranean Sea: a basin wide perspective, Ocean Science, 10(1), 69-92, 2014.

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

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.

Bergametti, G., Remoudaki, E., Losno, R., Steiner, E., Chatenet, B.: Source, transport and deposition of atmospheric Phosphorus over the northwestern Mediterranean, J. Atmos. Chem.. 14, 501–513, 1992.

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.

Bittig, H. C., Körtzinger, A., Neill, C., van Ooijen, E., Plant, J. N., Hahn, J., Johnson, K. S., Yang, B., Emerson, S. R.: Oxygen Optode Sensors: Principle, Characterization, Calibration, and Application in the Ocean. Frontiers in Marine Science 4, DOI: 10.3389/fmars.2017.00429, 2018.

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/s10236-017-1087-7, 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.

Copin-Montegut C.: Alkalinity and carbon budgets in the Mediterranean Sea. Global Biogeochemical Cycles, 7(4), 915-925, 1993.

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

Cornell, S., Rendell, A., and Jickells, T.: Atmospheric inputs of dissolved organic Nitrogen to the oceans, Nature, 376, 243–246, 1995.

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.

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 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.

de la Paz, M., Huertas, E.M., Padín, X.-A., Gónzalez-Dávila, M., Santana-Casiano, M., Forja, J.M., Orbi, A., Pérez, F.F., Ríos, A.F.: Reconstruction of the seasonal cycle of air–sea CO2 fluxes in the Strait of Gibraltar. Marine Chemistry, 126(1–4), 155-162, 2011.

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.

Estrada, M., Latasa, M., Emelianov, M., Gutiérrez-Rodríguez, A., Fernández-Castro, B., Isern-Fontanet, J., Mouriño-Carballido, B., Salat, J. and Vidal, M.: Seasonal and mesoscale variability of primary production in the deep winter-mixing region of the NW Mediterranean. Deep Sea Research Part I: Oceanographic Research Papers, 94, pp.45-61, 2014.

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, 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, 2009.

Guerzoni, S., Chester, R., Dulac, F., Herut, B., Loÿe-Pilot, M.-D., Measures, C., Migon, C., Molinaroli, E., Moulin, C., Rossini, P., Saydam, C., Soudine, A., Ziveri, P.: The role of atmospheric deposition in the biogeochemistry of the Mediterranean Sea. Prog. Oceanogr., 44 (1-3), 147-190, 1999.

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

Guyennon, A., Baklouti, M., Diaz, F., Palmieri, J., Beuvier, J., Lebaupin-Brossier, C., Arsouze, T., Béranger, K., Dutay, J.-C., and Moutin, T.: New insights into the organic carbon export in the Mediterranean Sea from 3-D modeling, Biogeosciences, 12, 7025-7046, https://doi.org/10.5194/bg-12-7025-2015, 2015.

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 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.

Herut, B. and Krom, M.: Atmospheric input of nutrients and dust to the SE Mediterranean, in: The Impact of Desert Dust Across the Mediterranean, edited by: Guerzoni, S. and Chester, R., Kluwer Acad., Norwell, Mass., 349–358, 1996.

Huertas, I. E., Ríos, A. F., García-Lafuente, J., Makaoui, A., Rodríguez-Gálvez, S., Sánchez-Román, A., Orbi, A., Ruíz, J., and Pérez, F. F.: Anthropogenic and natural CO2 exchange through the Strait of Gibraltar. Biogeosciences, 6, 647-662, 2009.

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.

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.

Johnson, K.S. and Claustre, H.: Bringing biogeochemistry into the Argo age, Eos, 97, 2016.

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.

Lévy, M., Resplandy, L., Klein, P., Capet, X., Iovino, D., & Éthé, C.: Grid degradation of submesoscale resolving ocean models: Benefits for offline passive tracer transport. Ocean Modelling, 48, 1-9, 2012.

Loÿe-Pilot, M. D., J. M. Martin, and J. Morelli: Atmospheric input of inorganic nitrogen to the western Mediterranean. Biogeochem., 9, 117-134, 1990.

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

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.

Manca, B., Burca, M., Giorgetti, A., Coatanoan, C., Garcia, M.J., Iona, A.: Physical and biochemical averaged vertical profiles in the Mediterranean regions: an important tool to trace the climatology of water masses and to validate incoming data from operational oceanography, J. Mar. Syst. 48, 83–116, https://doi.org/10.1016/j.jmarsys.2003.11.025, 2004.

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.

Mattia, G., Zavatarelli, M., Vichi, M., and Oddo, P.: The eastern Mediterranean Sea biogeochemical dynamics in the 1990s: A numerical study. Journal of Geophysical Research: Oceans, 118(4):2231-2248, 2013.

Mayot, N., D'Ortenzio, F., Taillandier, V., Prieur, L., Pasqueron de De Fommervault, O., Claustre, H., Bosse, A., Testor, P. and Conan, P.: Physical and biogeochemical controls of the phytoplankton blooms in North Western Mediterranean Sea: A multiplatform approach over a complete annual cycle (2012–2013 DEWEX experiment). Journal of Geophysical Research: Oceans, 122(12), pp.9999-10019, 2017.

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., https://doi.org/10.1029/2018GL080541, 2019.

Oddo, P., Adani, M., Pinardi, N., Fratianni, C., Tonani, M., and Pettenuzzo, D.: A Nested Atlantic-Mediterranean Sea General Circulation Model for Operational Forecasting, Ocean 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.

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.

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., 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.

Petihakis, G., Perivoliotis, L., Korres, G., Ballas, D., Frangoulis, C., Pagonis, P., Ntoumas, M., Pettas, M., Chalkiopoulos, A., Sotiropoulou, M., Bekiari, M., Kalampokis, A., Ravdas, M., Bourma, E., Christodoulaki, S., Zacharioudaki, A., Kassis, D., Potiris, E., Triantafyllou, G., Tsiaras, K., Krasakopoulou, E., Velanas, S., and Zisis, N.: An integrated open-coastal biogeochemistry, ecosystem and biodiversity observatory of the eastern Mediterranean – the Cretan Sea component of the POSEIDON system, Ocean Sci., 14, 1223-1245. https://doi.org/10.5194/os-14-1223-2018, 2018.

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 Eight EuroGOOS International Conference, Bergen, Norway, 3-5 October 2017, 335-342, 2018.

Raghukumar, K., Edwards, C.A, Goebel, N.L., Broquet, G., Veneziani, M., Moore, A.M., Zehr, J.P.: Impact of assimilating physical oceanographic data on modeled ecosystem dynamics in the California Current System. Prog. in Ocean., 138 B, 546-558, https://doi.org/10.1016/j.pocean.2015.01.004, 2015.

Ribera d'Alcalà M., Civitarese G., Conversano F., Lavezza R.: Nutrient ratios and fluxes hint at overlooked processes in the Mediterranean Sea, Journal of Geophysical Research, 108(C9), 8106, doi:10.1029/2002JC001650, 2003.

Richon, C., Dutay, J.-C., Dulac, F., Wang, R., Balkanski, Y., Nabat, P., Aumont, O., Desboeufs, K., Laurent, B., Guieu, C., Raimbault, P. and Beuvier, J.: Modeling the impacts of atmospheric deposition of nitrogen and desert dust-derived phosphorus on nutrients and biological budgets of the Mediterranean Sea. Prog. Ocean. 163, 21 – 39, https://doi.org/10.1016/j.pocean.2017.04.009, 2017.

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.

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 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.

Severin, T., Kessouri, F., Rembauville, M., Sánchez-Pérez, E.D., Oriol, L., Caparros, J., Pujo-Pay, M., Ghiglione, J.F., D'Ortenzio, F., Taillandier, V. and Mayot, N.: Open-ocean convection process: A driver of the winter nutrient supply and the spring phytoplankton distribution in the Northwestern Mediterranean Sea. Journal of Geophysical Research: Oceans, 122(6), pp.4587-4601, 2017.

She, J., Allen, I., Buch, E., Crise, A., Johannessen, J. A., Le Traon, P.-Y., Lips, U., Nolan, G., and other 5 co-authors: 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.

Siokou-Frangou, I., Christaki, U., Mazzocchi, M. G., Montresor, M., Ribera d'Alcalá, M., Vaqué, D., and Zingone, A.: Plankton in the open Mediterranean Sea: a review, Biogeosciences, 7, 1543–1586, doi:10.5194/bg-7-1543-2010, 2010.

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.

Stow, C.A., Jolliff, J., McGillicuddy, Jr D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., Rose, K.A., Wallhead, P.: Skill assessment for coupled biological/physical models of marine systems. J Marine Syst. 76(1–2):4–15. doi:http://dx.doi.org/ 10.1016/j.jmarsys.2008.03.011, 2009.

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.

Tanhua, T., Hainbucher, D., Schroeder, K., Cardin, V., Álvarez, M., and Civitarese, G.: The Mediterranean Sea system: a review and an introduction to the special issue, Ocean Sci., 9, 789-803, https://doi.org/10.5194/os-9-789-2013, 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, 2018.

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, https://doi.org/10.1016/j.cageo.2019.01.003, 2019.

Thierry V., Bittig H., The Argo-Bgc Team: Argo quality control manual for dissolved oxygen concentration. https://doi.org/10.13155/46542, 2018.

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.

Tsiaras, K. P., Hoteit, I., Kalaroni, S., Petihakis, G., and Triantafyllou, G.: A hybrid ensemble-OI Kalman filter for efficient data assimilation into a 3-D biogeochemical model of the Mediterranean. Ocean Dynamics, 67(6):673-690, 2017.

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, 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.

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

Yu, L., Fennel, K., Bertino, L., El Gharamti, M., and Thompson, K. R.: Insights on multivariate updates of physical and biogeochemical ocean variables using an Ensemble Kalman Filter and an idealized model of upwelling, Ocean Model., 126, 13-28, https://doi.org/10.1016/j.ocemod.2018.04.005, 2018.

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.

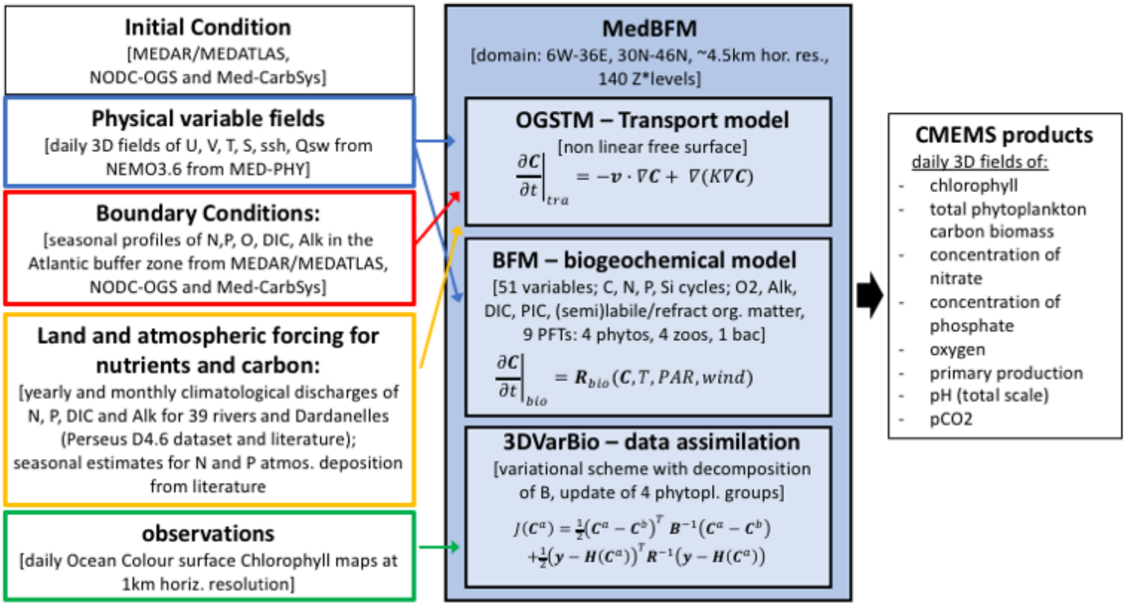

**Figure 1: The MedBFMv2.1 model system and interfaces with other components of CMEMS and external forcing data.**

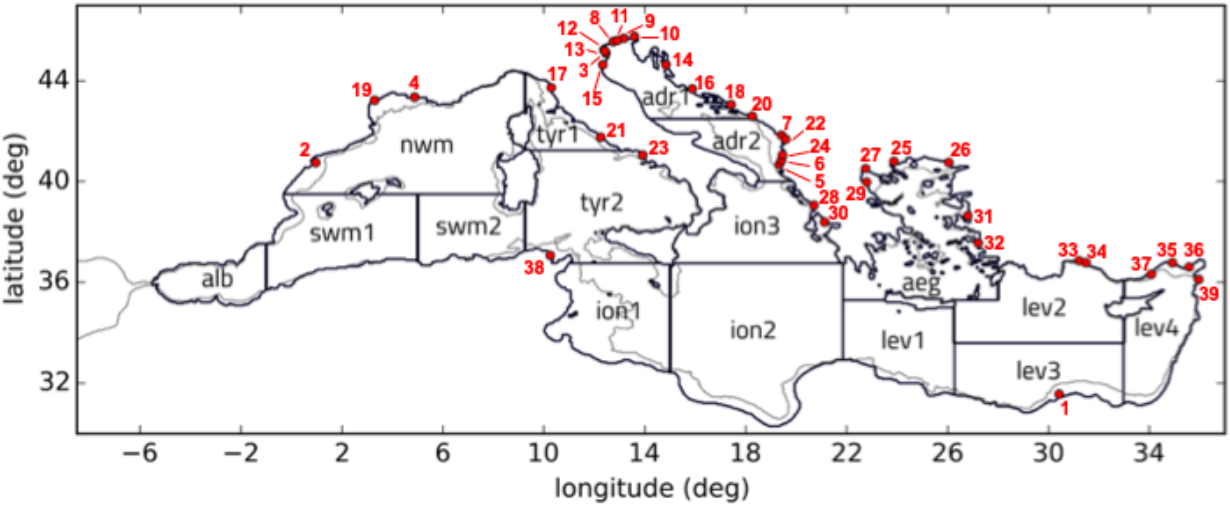

**Figure 2: Subdivision of the model domain in sub-basins used for the validation of the qualification run. According to data availability and to ensure consistency and robustness of the metrics, different subsets of the sub-basins or some combinations among them can be used for the different metrics: 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).**

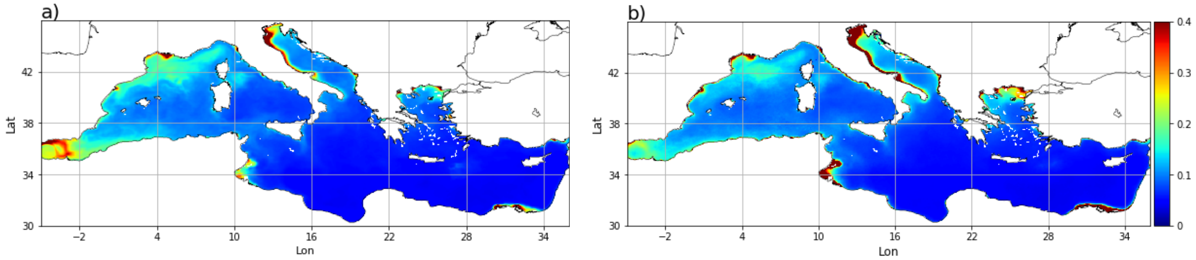

**Figure 3: Averaged annual maps of surface chlorophyll (mg/m3) from 2016-2017 qualification run (a) and from NRT multi-sensor satellite (b).**

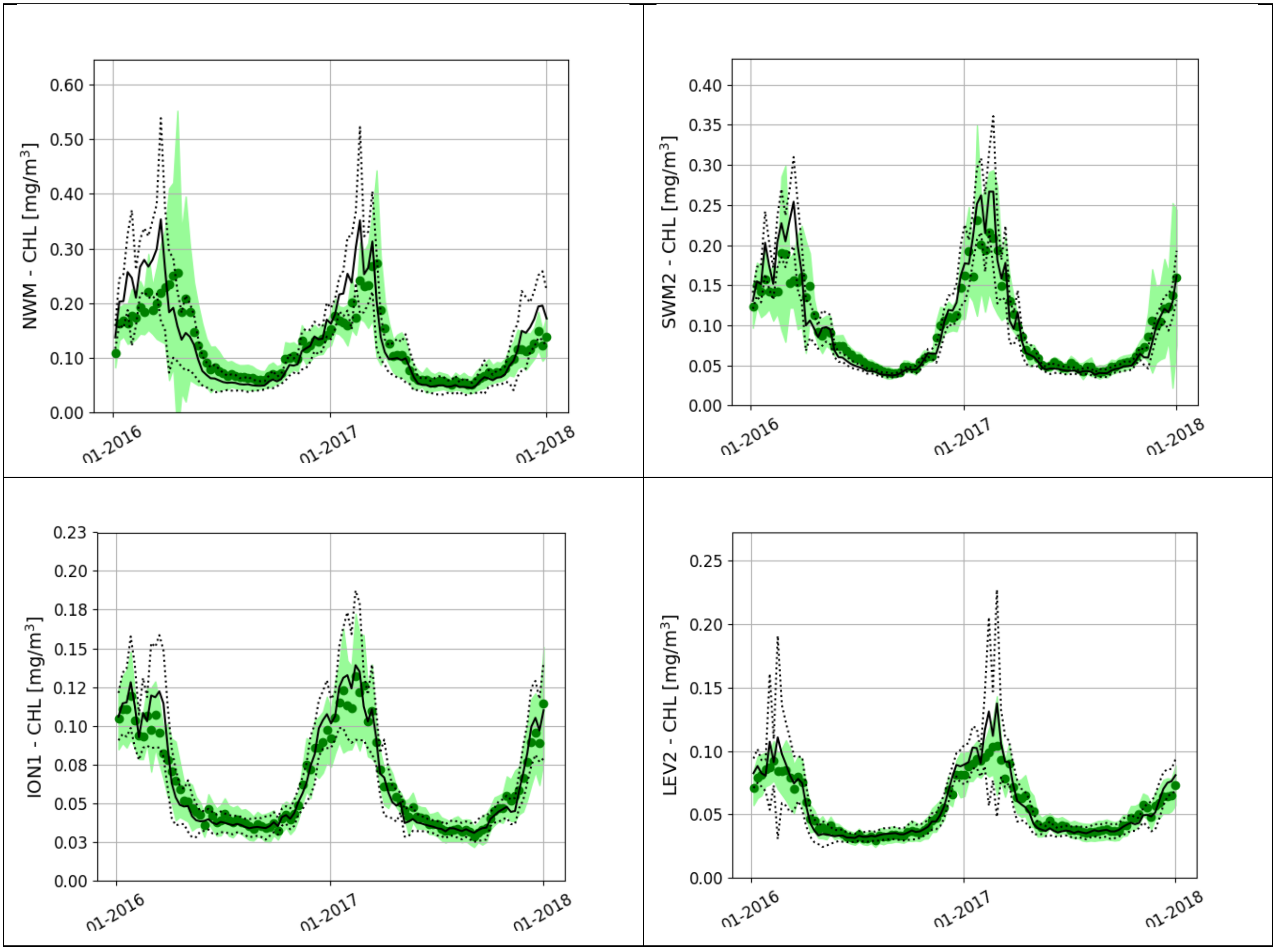

**Figure 4: Model (black line, with standard deviation in black dots) and satellite (green dots, with standard deviation covering the light green area) time series of mean surface chlorophyll concentration in open sea areas in four selected sub-basins of Fig. 2.**

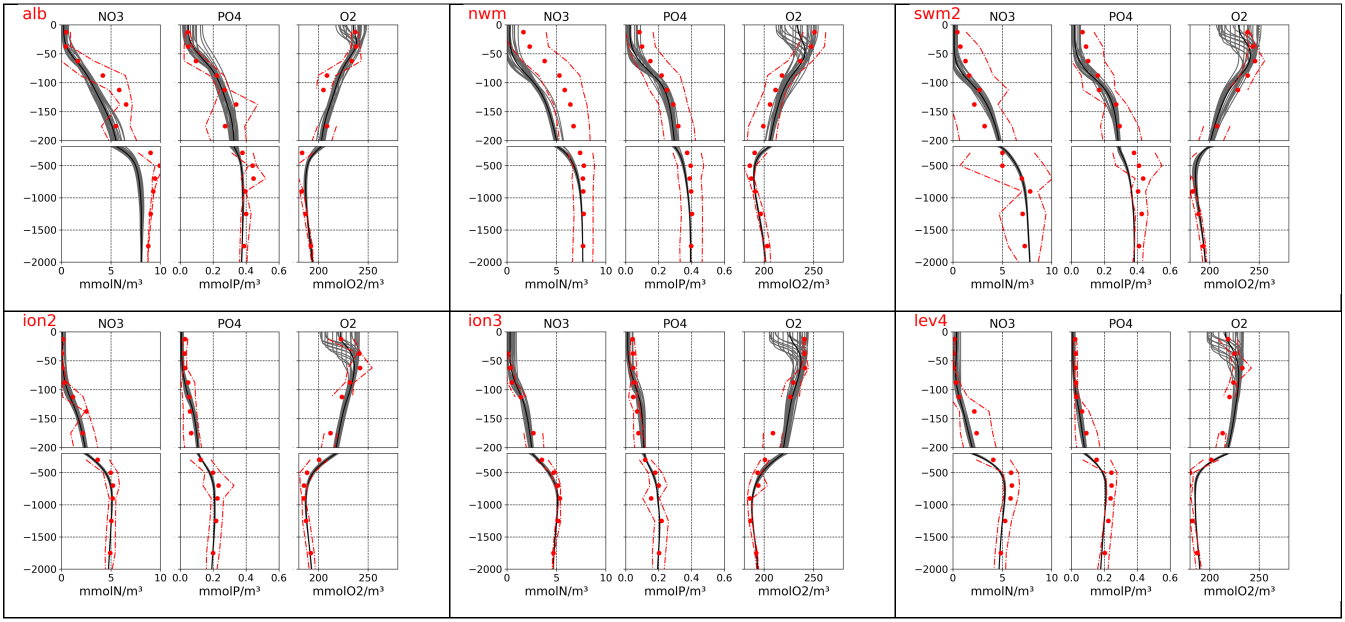

**Figure 5: Monthly (grey lines) and mean (black lines) vertical profiles from the qualification run for selected sub-basins of Fig. 2 compared with climatological profiles (red dots) and variability ranges (one standard deviation, red lines) of nitrate, phosphate and dissolved oxygen retrieved from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

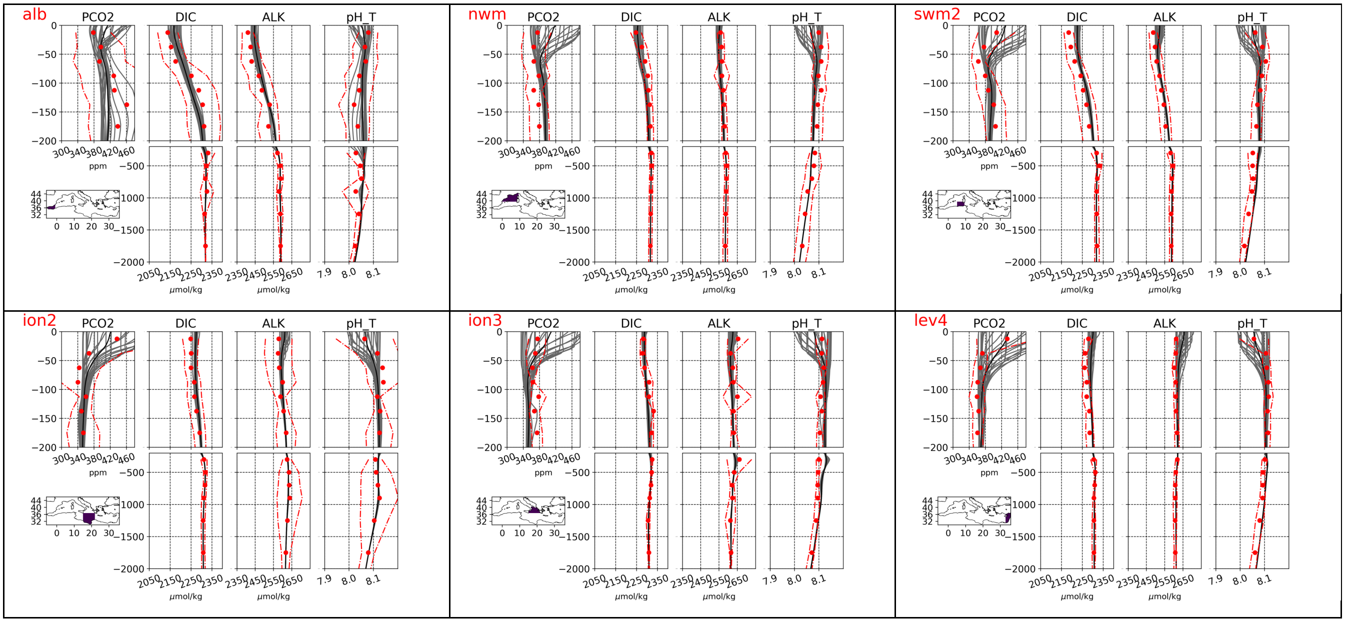

**Figure 6: Monthly (grey lines) and mean (black lines) vertical profiles from the qualification run for selected sub-basins of Fig. 2 compared with climatological profiles (red dots) and variability ranges (one standard deviation, red lines) of Dissolved Inorganic Carbon (DIC), alkalinity (ALK), pH 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).**

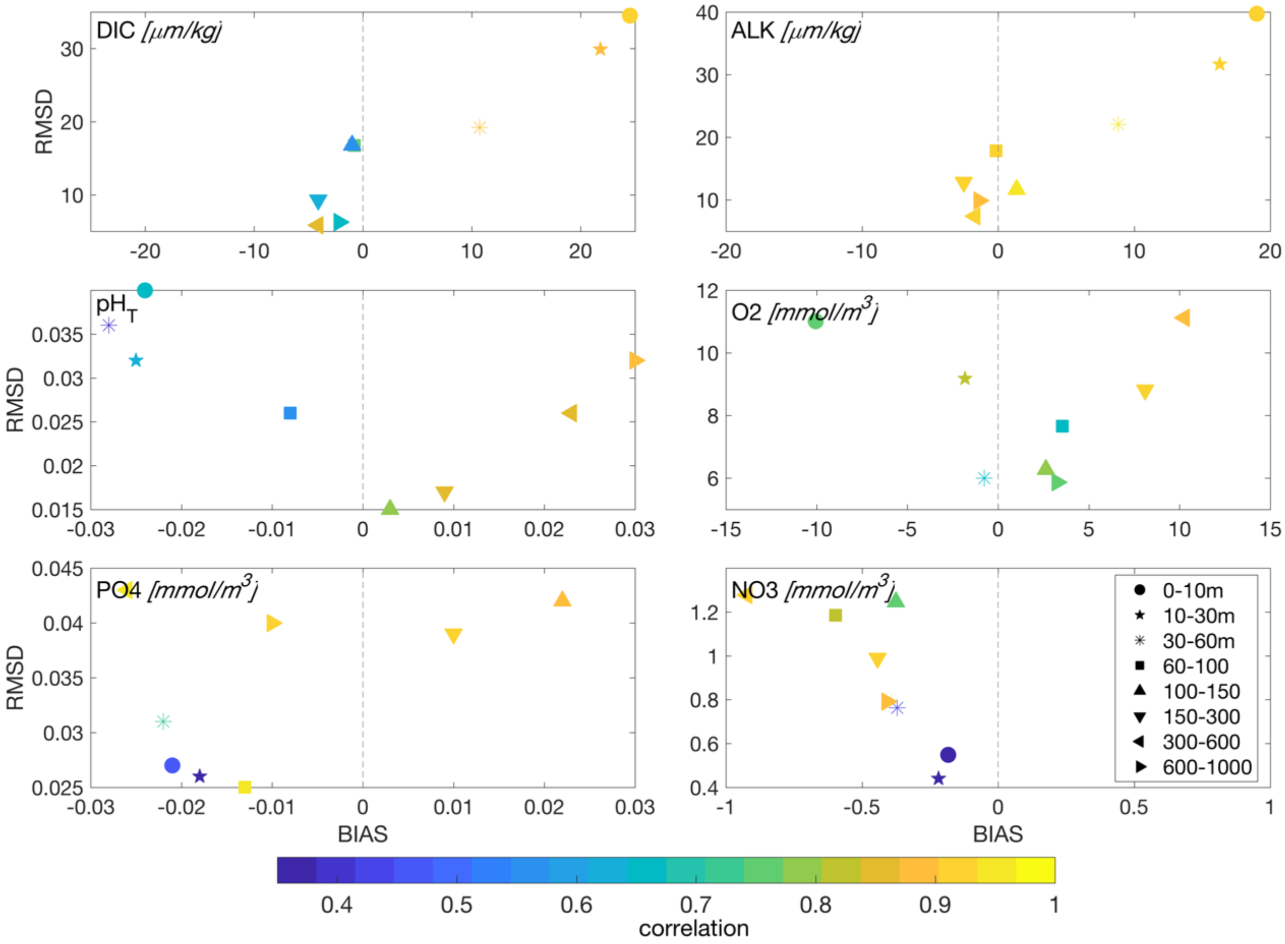

**Figure 7: Target diagrams and correlation (values in colour shading) between model and climatology for the different layers (symbols) and for the selected variables: alkalinity (ALK), dissolved inorganic carbon (DIC), oxygen (OXY), phosphate (PO4),**
5 **nitrate (NO3) and pH reported in total scale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

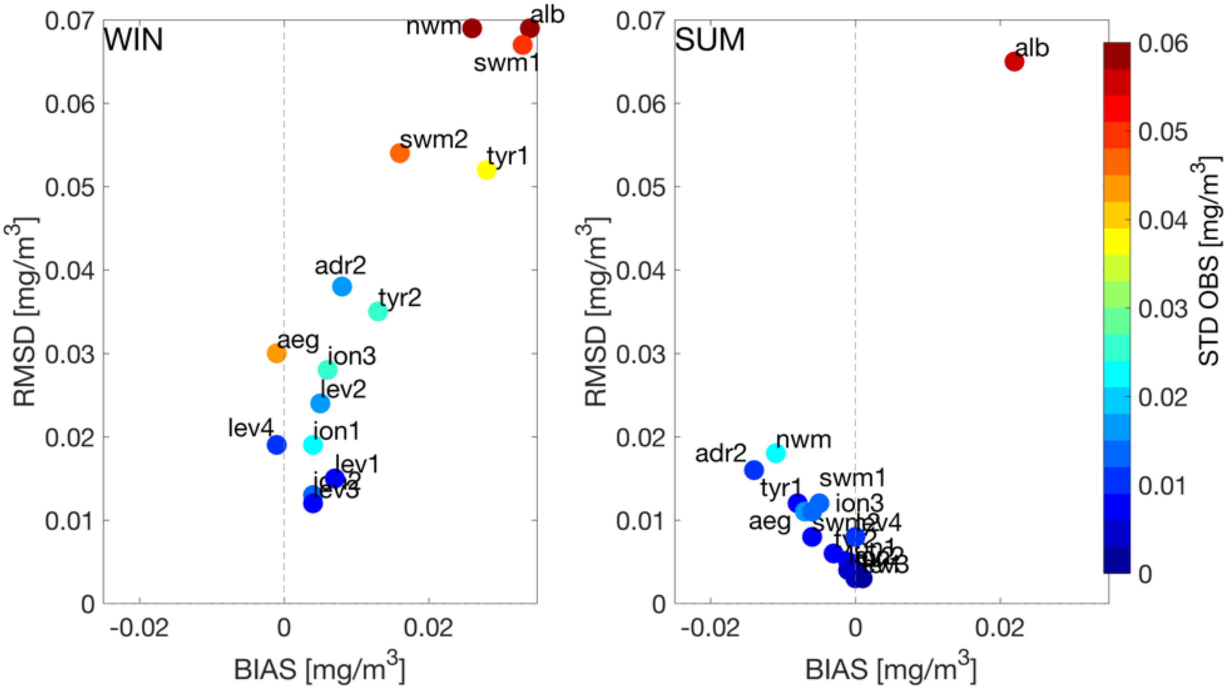

**Figure 8: Target diagrams of the model and satellite chlorophyll comparison and standard deviation of observations (in colour shading) for two periods: January to April (WIN, left) and June to September (SUM, right). For sake of readability, an offset in the values in WIN of [RMSD, BIAS] for ALB and 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.**

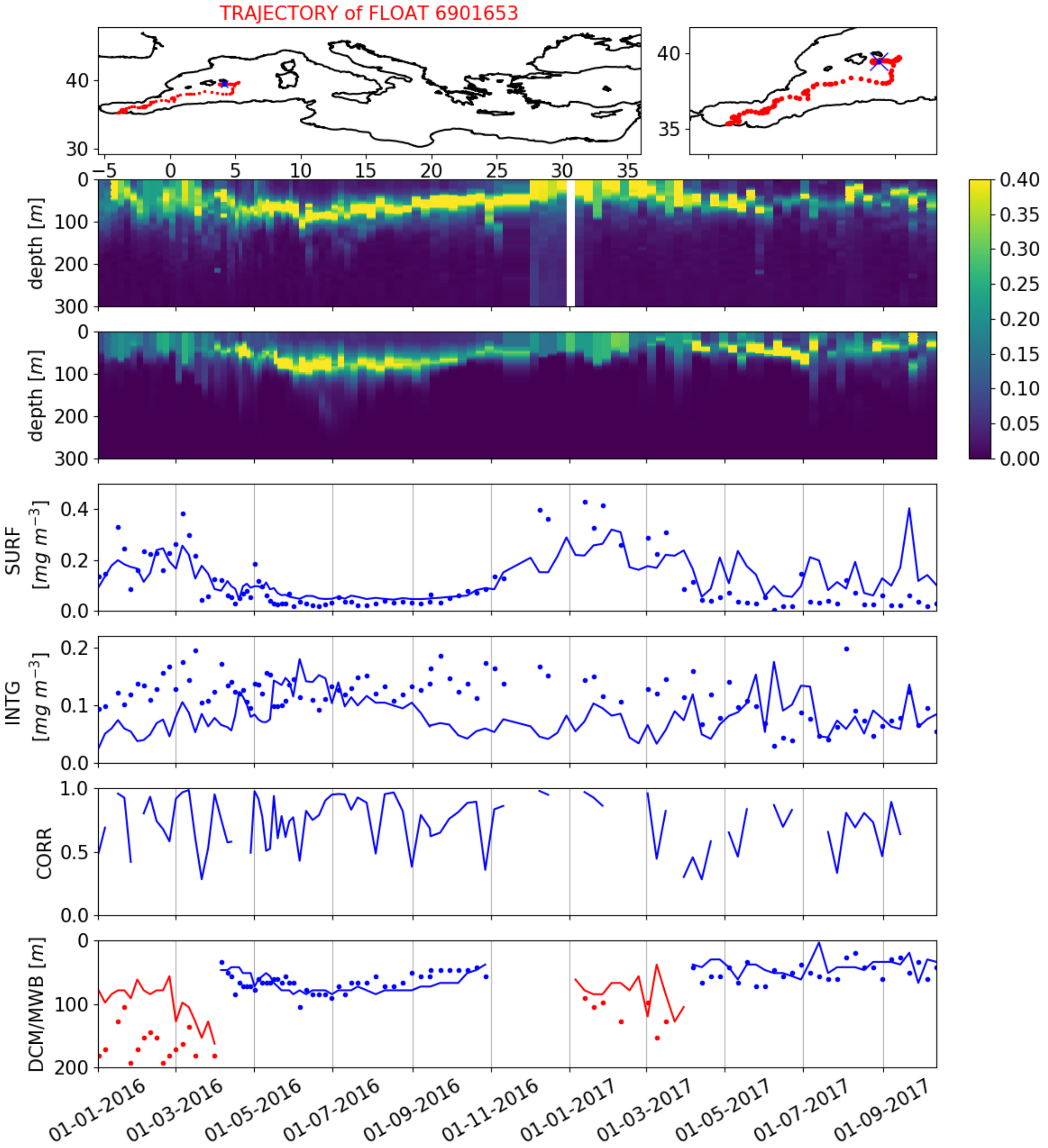

**Figure 9: Time evolution of BGC-Argo float 6901653. Top panel: trajectory of the BGC-Argo float (red dots), with deployment position (blue cross); Hovmoller diagrams of chlorophyll concentration (mg/m$^3$) from float data (2nd panel) and model outputs (3rd panel) matched-up with float position for the period 2016-2017. Computation of selected skill indexes for model (solid line) and float data (dots): surface chlorophyll (SURF, 4th panel) and 0-200m vertically averaged chlorophyll (INTG, 5th panel), correlation between vertical profiles (CORR, 6th panel), depth of the deep chlorophyll maximum (DCM, blue) and depth of the mixed layer bloom in winter (MWB, red; bottom panel).**

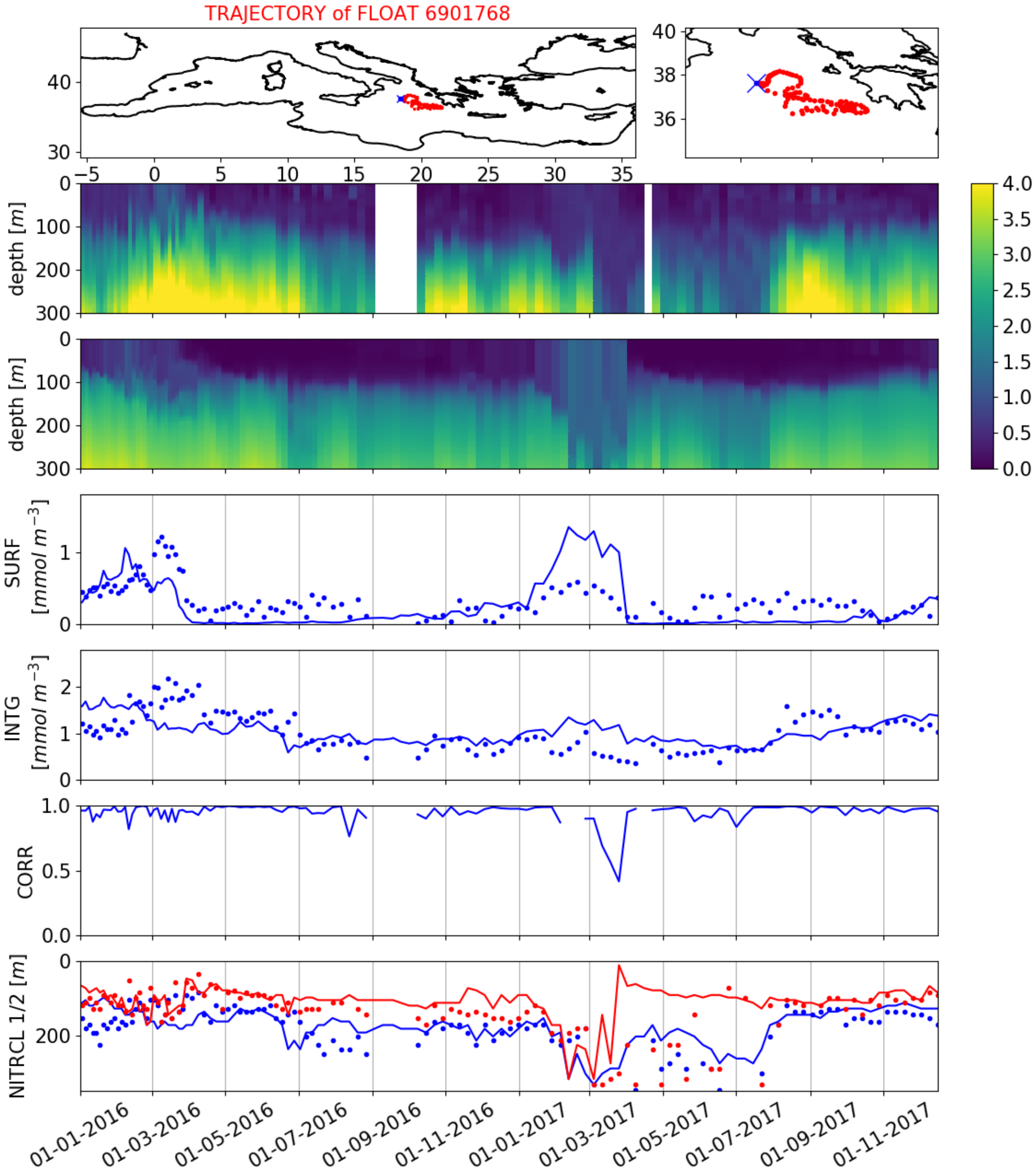

**Figure 10: Time evolution of BGC-Argo float 6901768. Top panel: trajectory of the BGC-Argo float (red dots), with deployment position (blue cross); Hovmoller diagrams of nitrate concentration (mmol/$m^3$) from float data (2nd panel) and model outputs (3rd panel) matched-up with float position for the period 2016-2017. Computation of selected skill indexes for model (solid line) and float data (dots): nitrate concentration at surface (SURF, 4th panel) and 0-200 m vertically averaged concentration (INTG, 5th panel), correlation between vertical profiles (CORR, 6th panel), depth of the nitracline computed as NITRCL1 (blue) and NITRCL2 (red; bottom panel).** 5

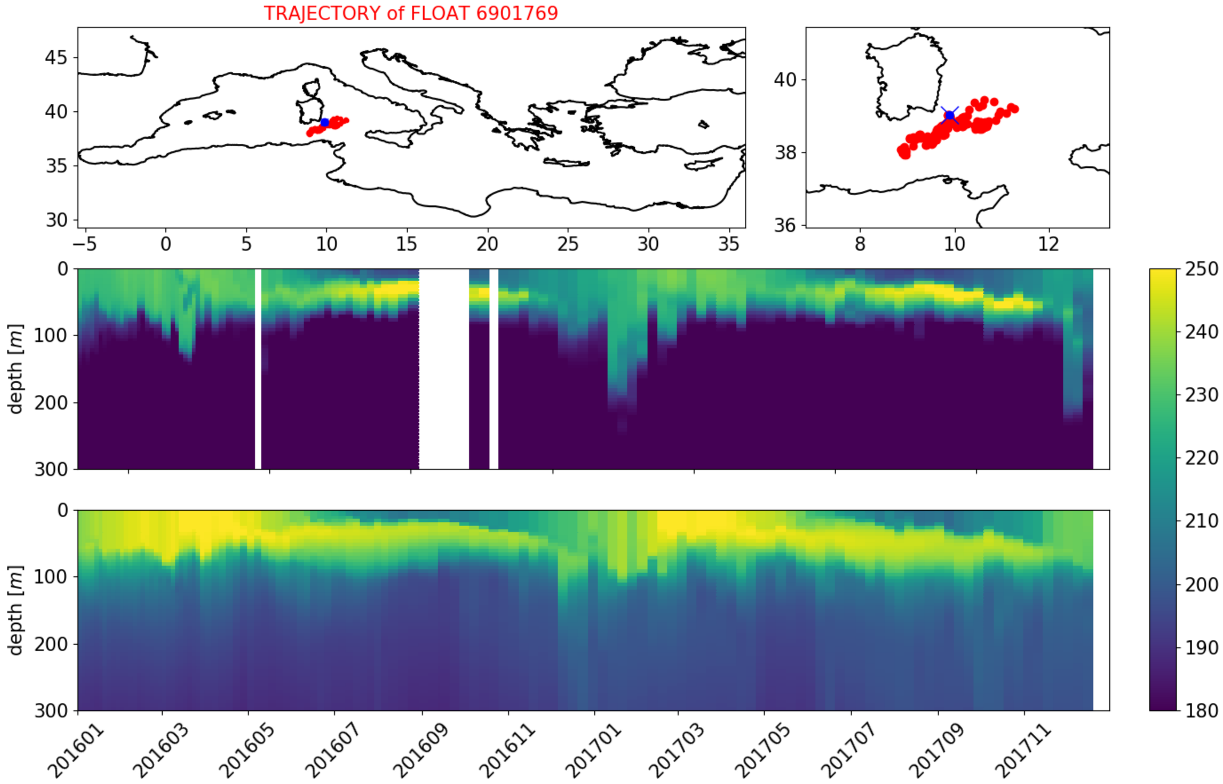

**Figure 11: Time evolution of BGC-Argo float 6901769. Top panel: trajectory of the BGC-Argo float (red dots), with deployment position (blue cross); Hovmoller diagrams of oxygen concentration (mmol/m$^3$) of one selected BGC-Argo float 6901769 (middle panel) and model outputs (bottom panel) matched-up with float position for the period 2016-2017.**

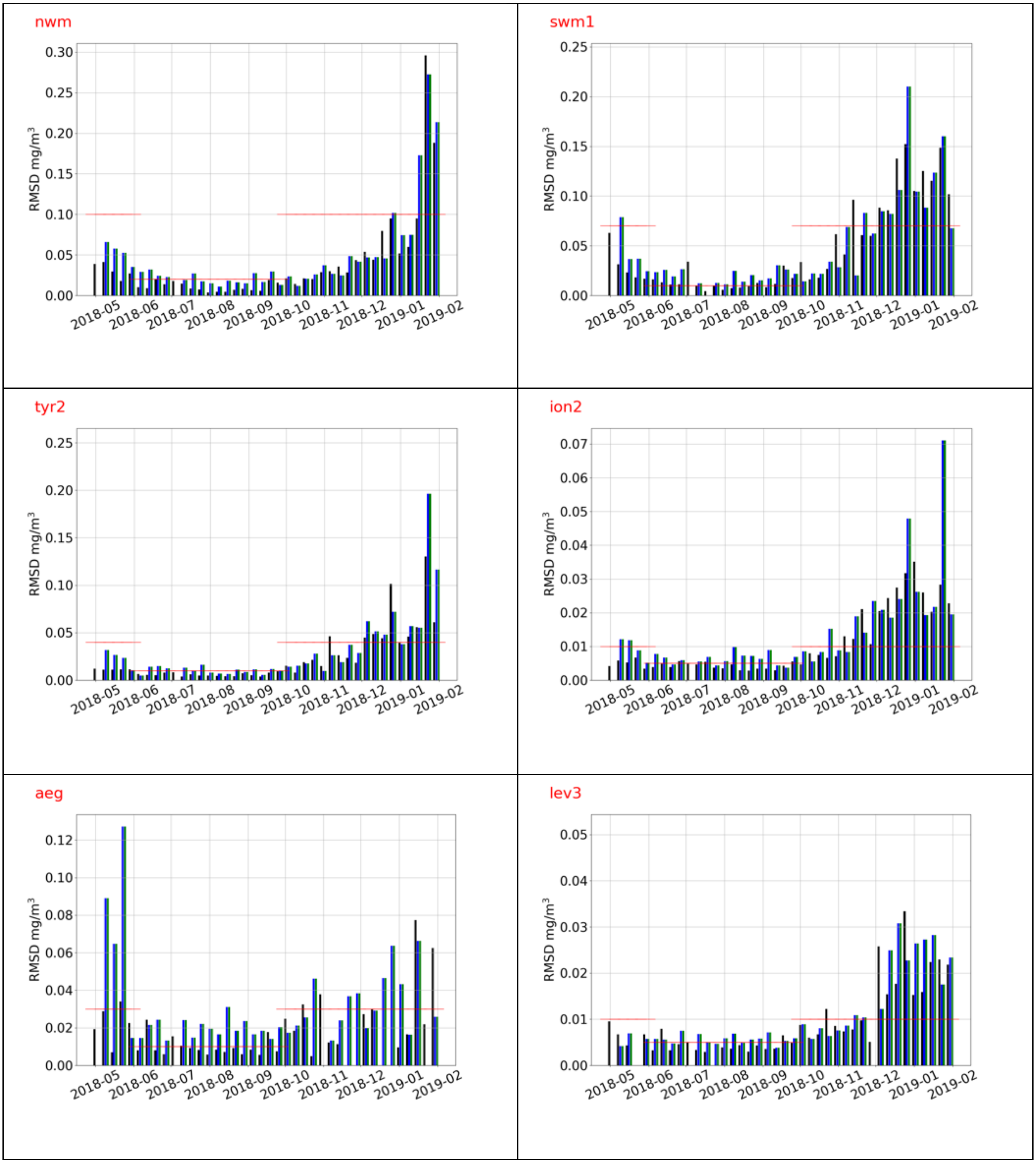

**Figure 12: Sub-basin RMSD between surface chlorophyll model forecast at lead time 24 (T1, black), 48 (T2, blue) and 72 hours (T3, green) and daily satellite maps. As benchmark reference, the two seasonal mean RMSD values computed from 2016-2017 pre-operational qualification run are shown (red lines).**

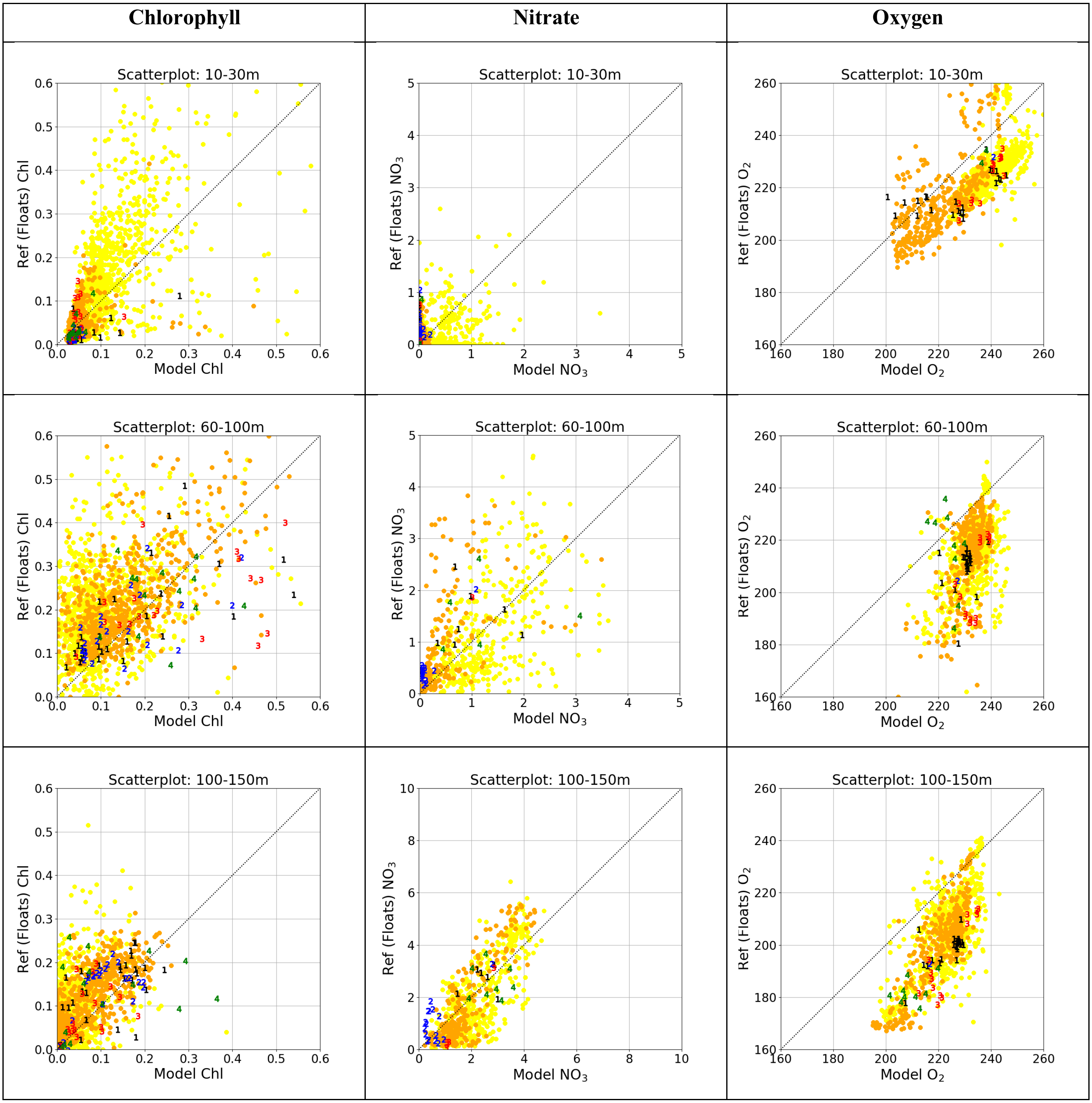

**Figure 13: Scatter plots of reference (y-axis) versus model forecast (x-axis) for chlorophyll (left column), nitrate (middle column) and oxygen (right column) at different vertical layers: 10-30 m, 60-100 m and 100-150 m. Model forecast are labelled with numbers from 1 to 4 corresponding to lead time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

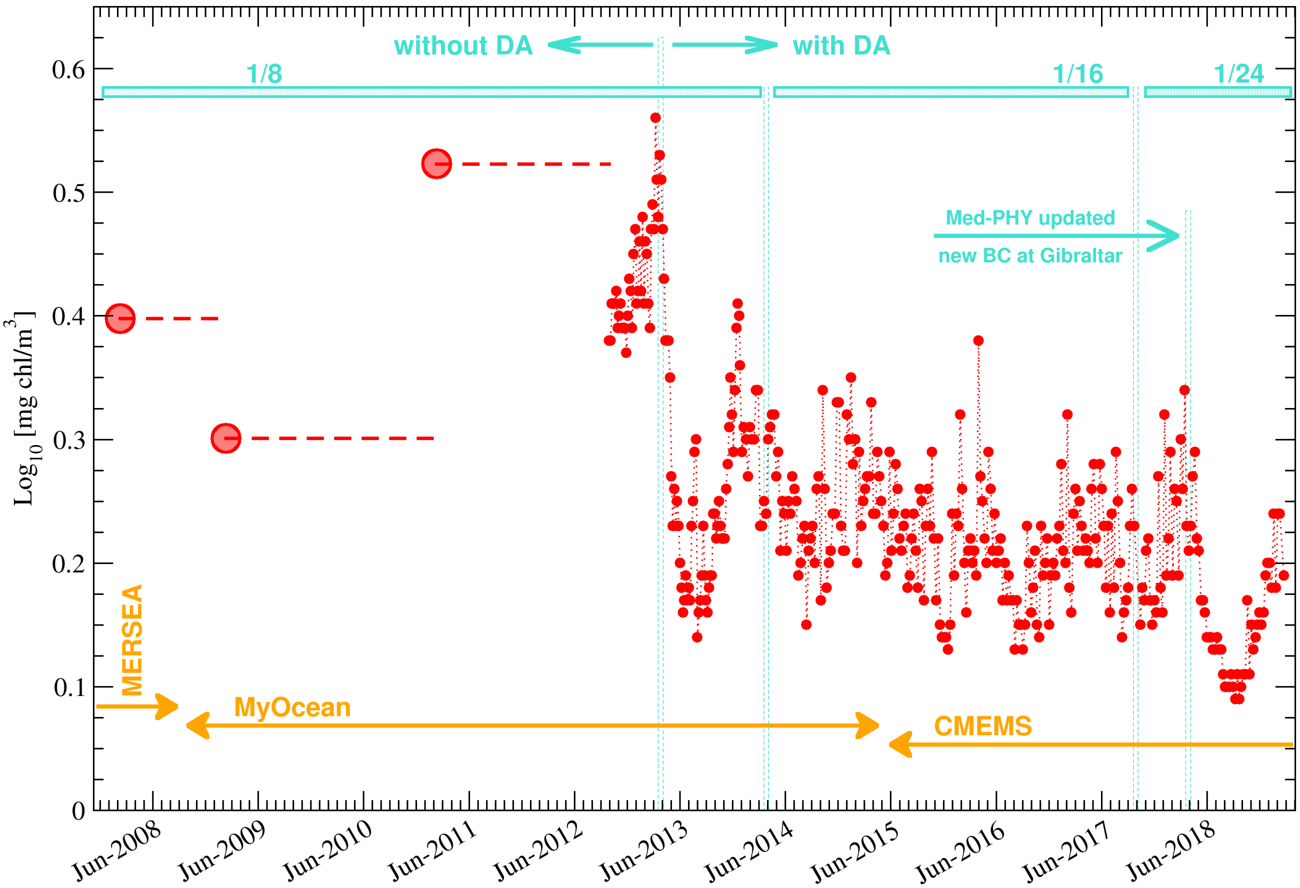

**Figure 14: RMSD of the surface chlorophyll concentration with satellite data. The use of logarithmic units has been the standard for RMSD since the implementation phase. Regular, weekly product quality assessment (red dots and dotted line) started at the end of 2012. Before, quality assessment was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

| | 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/m$^2$/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;<br>86-232*;<br>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 |

**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/m$^3$] | | INTG [mg/m$^3$] | | | 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 |

**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/$m^3$] |  | INTG [mmol/$m^3$] |  | 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, for the period January 2016 – December 2017.**

|  | PRE-OPERATIONAL |  | T1 | T2 | T3 | T4 |
|---|---|---|---|---|---|---|
| **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 qualification run and for the first 4 days of forecast of the Med-BIO forecast system (T1 to T4) since April 2018. Statistics are computed using the layers 0-300 m for nitrate and oxygen and 0-150 m for chlorophyll.**