Dear editor,

we considered all the proposed suggestions for the manuscript's correction. Below we reply to specific questions and uncertainties arising from the previous version of the text. We decided to remove some plots (Figs. 12, 13 and 15 from the first revision's version) without affecting the message of the manuscript and at the same time reducing its' length. The plot related to sensitivity analysis were moved in the Supplementary material.

Pg.2, c1: There seems to be a bit of a contradiction here: if radiometric data is more accurate, how does it help to use it to estimate chlorophyll values that are then validated using the supposedly "inaccurate" BGC-Argo fluorescence-derived Chlorophyll?

Measured Chl data could be inaccurate in terms of magnitude, but not in terms of DCM depth, even though it's derived from fluorescence. The validation is therefore more concentrated on the DCM depth rather than magnitude due to the uncertainties associated with the latter. We would keep the sentence as it is.

Pg.2, c4: Why is there is no mention of the DCM. Isn't that your most important validation parameter?

We agree with the editor and modified the paragraph accordingly.

Pg.5, c1: I am not sure I understand the meaning of this sentence.

It's indeed redundant within the scope of the present study, so we cancelled it.

Pg.6, c1: Would it be possible to include the depth range or intervals from which the parametrizations are derived (or give the info in Table 2).

z_{max}in Table 2 is already the depth range. The regression analysis is carried out for mean values of depth layers of 15 m thickness, which has been specified also in Sect. 2.2.2.
We modified the Table 2 caption accordingly.

Pg.6, c4: Shouldn't the units be 1m/s?

Eddy diffusivity coefficients are expressed in m^2/s.

Pg.9, c1: This section title is somewhat misleading. This section covers all of 22 pages of the manuscript and does not only describe the reference simulations but also sensitivity analysis of the REF model which is not announced in the introduction and methods.

According to the editor's suggestion, we moved the paragraph related to the sensitivity analysis to the supplementary material.

Pg.9, c2: What do the legend (names and colors) represent? I assume data from individual floats? over what time range each?
Each color represents one BGC-Argo float. The chart reporting the WMO code of each BGC-Argo float is in the Supplementary material. We updated it by adding the float tracking period information.
What is the p-value for the regression? (analysis of variance or t-test?) are the correlations statistically significant?
t-test. Yes, the correlations are statistically significant. p-value < 0.005 (as reported in the caption of Fig.2)

Pg.10, c3: Do the authors mean at the surface of the ocean, or do they mean something like "primarily" or "overall"?
At the surface of the ocean, we modified the sentence accordingly.

Pg.10, c5: Can you please explain what the "initial conditions statistics" mean?
We evaluated the initial conditions based on reanalysis (used in the 3-dimensional model configuration) versus BGC-Argo float (Chl) data. This was additionally tested to make sure that the initial conditions of our 1-dimensional model do not degrade when integrated in time. Since it's not essential information we deleted this sentence.
Further here R is in capital letters but not in the figure 2 for example.
We corrected this as well.

Pg.15, c2: I am not sure I understand the link between lateral advection and strong vertical gradients in nutrient inventories. Can this be clarified?
Both lateral advection and vertical mixing could impact the nutrient inventory variability, but here we verify that the most important process is vertical mixing (regarding DCM depth features). As specified in the text, data driven mixing and vertical turbulence effects allow to simulate correctly the seasonal variability of the DCM depth.
We modified the paragraph, hopefully making it easier to understand.

Pg.15, c9: What is the p-value for the regression? (analysis of variance or t-test?) are the correlations statistically significant?
See reply for Pg.9, c2.

Pg.15, c11: It would be helpful if the same units were used in all graphs (see figs. 3-6 where PAR is given as µmol/m2/s.
The units are now uniform in all plots, i.e. we chose **µmol/m2/s**

Pg.17, c1: This whole section could go in an appendix as these are additional experiments not mentioned in the introduction. See also previous comments.
We followed the suggestion and moved the section in the Supplementary material.

Pg.18, c1: Why is the CL1 simulation shown here. Shouldn't this be shown in the next sections?
See reply above (pg.17, c1)

What is the p value for the regression? (analysis of variance or t-test?) are the correlations statistically significant?
See reply for Pg.9, c2.

Please explain how average Chl 0-25m was estimated as well as what dataset (which BGC-Argo float) was used here.
We calculated the mean value of Chl in the first 25 meters for all profiles when DCM was present. The same criterion was applied for both data and model.

Pg.18, c4: Thickness? Biomass? Please specify what magnitude means.
Biomass. We decided to remove this picture as mentioned in the beginning of the reply.

Pg.19, c8: "indicate that surface"
We don't refer only to surface nutrients.

Pg.21, c1: Please explain how BIO and PO4 were estimated. Are those model results or field measurements?
The sentence was modified in order to make it clear that we are talking about model results and not field measurements. There are no data available for phosphates from the BGC-Argo network.

Pg.22, c1: Why not present the full results from these simulations here? Do the values in text correspond to results shown in Fig. 14?
We prefer to keep only one summarizing plot to avoid having too many of them.

Pg.23, c1: Why is RMSD negative? Ed. 6 given in p.8 should give positive values only.
Yes we agree, we modified the figure by plotting only the positive axis.

Pg.24, c3: I fail to see that fig. 16 (16a in particular) shows a east-west gradient since only residuals are plotted.
We added another subplot with monthly climatology DCM for west and east separately. This should show the gradient described in the paper.

Pg.25, c1: The statement based on which data/simulation?
The statement is based on a reference we added also at the end of the sentence (Crispi et al., 2001).

Pg.29, c1: I have noticed differences between the REF and measured profiles in this figure compared to Fig. 15. Can this be explained? are those data for different time periods? For the sake of comparison wouldn't it make sense to choose the same time period?
They are the same, we checked again (both the red – REF - and dark blue – DATA - lines). The other curves are different because we consider different subsets of simulations.

Pg.30, c6: The meaning of this sentence is unclear.
We deleted it the second part of the sentence, which made it unclear.

Referee comment on the revised version

I acknowledge the effort made by the authors to comply with referees' suggestions in the revised version but I think that the paper needs a further effort to be ready for being published. This for the following reasons:

1. One of the declared scopes of the paper (e.g., Title, p.1 l.5, p1 l.23-p.2 l.7) was to explore the advantage of assimilating Argo profiles in a coupled model. Reading the manuscript I understood that the assimilated variables were, in turn, PAR and chlorophyll a profiles to compute PAR with several bio-optical models. As for the mixed layer depth, which is one term to modulate the diffusivity profile, it is not clear if it responds to external forcing, e.g., p.5 l.5, or is also an assimilated variable. I might have missed this. If not, this should be clarified.

As explained in Sect. 2.2.1, the mixed layer depth is computed from temperature and salinity measured from BGC-Argo floats and it's used in the mixing model.

2. On the other hand, a significant part of the text is devoted to discuss not too much the feasibility of ARGO data assimilation in a coupled model but the mechanisms determining the DCM dynamics. This is interesting but, apparently, the authors do not analyze the basic mechanism behind the functioning of the DCM. At a first order of approximation the DCM is the depth where the upward diffusive nutrient flux is fully uptaken by, prevalently, phytoplankton. This is why the isolume is a good, first order, proxy for DCM depth. This has been discussed by Letelier e al. (L&O *49*(2), 508-519, 2004) and, more recently by Cullen (AnnRevMarSci, 2015) none of whom is cited in the paper. Of course, there might be phylogenetic or ontogenetic adaptations, but I assume that the model has a constant physiology for phytoplankton. I would hypothesize that increasing vertical diffusion should certainly increase the carrying capacity, and therefore, the DCM amplitude, which is what the authors observe, but it should also move the DCM depth upwards to reach a new steady state where the nutrient flux is utilized at the higher rate because of higher photon flux. Diffusion does also disperse cells but the authors focus mostly on this aspect, i.e., the thickness of the DCM not on its depth dependence on diffusion and vertical gradients.

Yes, we evaluated this in a sensitivity analysis, analogous to the one reported in the 20x20 bivariate perturbations experiment (see Supplementary material). Perturbing mixing for 4 orders of magnitude (10^{-6} to 10^{-2} m^2/s), the average difference is around 10 meters in DCM depth, with higher DCM when mixing is higher, as mentioned above by the reviewer. We added suggested references with a brief discussion on the role of mixing.

3. The authors focus on phosphate as the possible driving nutrient. It may depend on the existing paradigm that phosphate is the 'limiting' nutrient in the Mediterranean Sea. It might be interesting to examine the nitrate behavior. However, the intriguing

pattern is that the phosphate concentration in the WMED is approximately double than in the EMED at the same isolume but the chlorophyll is more or less the same, for what can be seen from figures 3 to 6. How the authors interpret this, since the phytoplankton physiology should be the same? May be that the similarity is a bias of the graphic representation.

We agree with the reviewer's comments. In future we plan to consider also nitrate data from BGC-Argo floats, which could be an additional validation parameter.
The Chl content at DCM appears higher in the western basin, but the response is nonlinear: double nutrient concentration does not directly imply double Chl concentration.
In the sensitivity analysis section (Supplementary material) we evaluated the response in Chl concentration by further increasing the concentration of nutrients. The effect of increasing nutrients is also evaluated in terms of self-shading in the section related to bio-optical models.

4. Linked to the above is the sensitivity of the DCM depth to phosphate (nutrient?) profile. Swapping East and West Argo profilers the authors (see response) state that there is no significant effect. Indeed, the slope of the model DCM vs observed DCM depth shows that the model underestimates the DCM depth for deep DCM and slightly overestimates the depth for shallow DCM, a pattern that is not discussed. More important, when they swap the profiles the model enhances this feature, which I would interpret as the fact that the higher irradiance in the EMED produce shallower DCM than in the real environment and the opposite occurs in the WMED. The scatter plot in Figure 2 does not allow a simple geo-localization of the Argos but, in any case, I would not consider the result of the analysis as a demonstration that nutrient profile has a minor role in determining the DCM depth.

As explained in the conclusions section, we are not stating that the role of nutrients in shaping the DCM is absent. The evaluation of the role of nutrients in this manuscript is performed in two stages:
1) In the REF simulation only direct effects of nutrients are accounted for, e.g. in relation to vertical diffusion and the corresponding nutrient upward fluxes. In this case we observed that light appears to play a mayor role in shaping the gradient.
2) If we consider indirect effects of nutrients on light propagation, we see how nutrients play a role trough self-shading. This is demonstrated through the analyses which used alternative bio-optical models that account for self-shading effects.

5. More important, even not being an English mother language, I think that the text should be revised both in the wording and in the way the work done is presented. I still found some parts hard to follow and to connect to the others.
We followed the reviewer's suggestion and revised the text thoroughly.

In the following part of the document we attach a comparison of the first submitted manuscript with the presently submitted one (R2) processed with latex diff software in order to show the changes performed.

Merging bio-optical data from Biogeochemical-Argo floats and models in marine biogeochemistry

Elena Terzić[1,4], Paolo Lazzari[1], Emanuele Organelli[2], Cosimo Solidoro[1], Stefano Salon[1], Fabrizio D'Ortenzio[2], and Pascal Conan[3]

[1]Istituto Nazionale di Oceanografia e di Geofisica Sperimentale - OGS, Via Beirut 4, 34151 Trieste, Italy
[2]Sorbonne Universités, CNRS, Laboratoire d'Océanographie de Villefranche, LOV, F-06230, Villefranche-sur-Mer,France
[3]Sorbonne Université, Pierre et Marie Curie-Paris 06, CNRS - UMR7621 LOMIC, F66650 Banyuls-sur-Mer, France
[4]Università degli Studi di Trieste, Dipartimento di Matematica e Geoscienze, Via E. Weiss 2, 34128 Trieste, Italy

Correspondence to: Paolo Lazzari (plazzari@inogs.it)

Abstract. In numerical models for marine biogeochemistry, bio-optical data ~~, such as measurements of the light field,~~ may be important descriptors of the dynamics of primary producers and ultimately of oceanic carbon fluxes. However, the paucity of field observations has limited the integration of bio-optical data in such models so far. New autonomous robotic platforms for observing the ocean, i.e., Biogeochemical-Argo floats, have drastically increased the number of vertical profiles of irradiance, photosynthetically available radiation (PAR) and algal chlorophyll concentrations around the globe independently of the season. Such data may be therefore a fruitful resource to improve performances of numerical models for marine biogeochemistry. Here we present a work that integrates ~~into a 1-dimensional model~~ 1314 vertical profiles of PAR acquired by 31 BGC-Argo floats operated in the Mediterranean Sea between 2012 and 2016 into a 1-dimensional model to simulate the vertical and temporal variability of algal chlorophyll concentrations. In addition to PAR as input, alternative light and vertical mixing models were considered. We ~~evaluated~~ evaluate the models' skill to reproduce the spatial and temporal variability of deep chlorophyll maxima as observed by BGC-Argo floats. The assumptions used to set up the ~~1-D~~ 1-dimensional model are validated by the high number of co-located ~~in-situ~~ in situ measurements. Our results illustrate the key role of PAR and vertical mixing in shaping the vertical dynamics of primary ~~produces~~ producers in the Mediterranean Sea. Moreover, we demonstrate the importance of modeling the diel cycle to simulate chlorophyll concentrations in stratified waters at the surface.

1 Introduction

In most biogeochemical models ~~, which are successfully coupled with hydrodynamics,~~ the description of optics is generally (over)simplified~~, therefore one of the necessary improvements still remains the integration of a~~ . The integration of more complex optical ~~model~~models, where inherent and apparent optical properties (IOPs and AOPs respectively) are already included as model state variables, (**?**), constitutes one of the necessary improvements. The research community is emphasizing the importance of merging different methods in order to improve the skill of numerical models, such as the assimilation of remote sensing data or the use of ~~in-situ~~ in situ data both for initialization and validation purposes. Until recently, the use of the latter

was especially critical due the scarcity of observations, however the emergence of autonomous robotic platforms such Biogeochemical Argo floats (hereafter BGC-Argo) helped ~~filling~~ reducing the gap in bio-optical measurements acquired around the globe, regardless of the season.

~~In particular, the Mediterranean Sea Monitoring and Forecasting Centre (Med-MFC) operatively produces analyses, forecasts and reanalyses of a series of biogeochemical state variables (e.g. chlorophyll, nutrients, pCO2) for the European Copernicus Marine Environment Monitoring Services (CMEMS) since 2015 using the MedBFM model (???), which assimilates surface chlorophyll from satellite observations (??).~~The introduction of BGC-Argo floats has led to a drastic increase in the number of radiometric measurements in the Mediterranean Sea, such as downward irradiance (Ed) and photosynthetically available radiation (PAR), for which specifically developed quality control procedures and refined sensor calibration (**??**) have widespread their use (**????**). BGC-Argo can therefore be an important source of high vertical spatial and temporal resolution data that can be integrated in the calibration and tuning of bio-optical numerical models for understanding marine biogeochemistry processes. ~~To this end the Mediterranean Sea proves to be an important region to study due to its bio-optically anomalous nature. It is characterized by complex trophic gradients (???) and spatially heterogeneous inherent optical properties (?). Such gradients are mainly related to the inverse estuarine circulation of the area (?) and to the varying distribution of optically significant substances (e.g. colored dissolved organic matter - CDOM; non-algal particles - NAP) that modulate the light penetration along the water column (?). Moreover, inherent optical properties (IOPs) could be affected also by important processesof Saharan dust deposition (?).~~

At present, no studies have tried to assimilate radiometric quantities into numerical models to improve the simulation of chlorophyll dynamics in ~~this basin and~~ the Mediterranean Sea and to investigate the causes of the vertical, spatial and temporal variability ~~eastward. Assimilating radiometry could prove~~ of zonal gradients. Assimilating radiometric data could prove to be more robust than chlorophyll assimilation as a result of a more accurate uncertainty characterization of optical measurements (**??**) compared to other biogeochemical variables, such as fluorescence-derived chlorophyll.

Specific studies are required to demonstrate to what extent the assimilation of radiometric data can improve the model skill in simulating key biogeochemical variables (e.g. chlorophyll, nutrients, primary productivity). In this paper we develop a 1-dimensional (1-D) model that assimilates PAR profiles acquired by BGC-Argo floats in order to replicate the vertical and temporal dynamics of phytoplankton chlorophyll concentrations. As a first modelling attempt, the exploration is carried out with a "voxel" approach, where light and mixing conditions were replicated from data available from floats. We analyse and validate model performances through a comparison of model outputs with the high number of co-located vertical profiles of chlorophyll concentrations (derived from fluorescence) measured by BGC-Argo floats. In particular, such kind of analysis allows to study some of the drivers modulating the DCM depth and amplitude in stratified conditions. Subsequently, we test different mixing and bio-optical models that simulate downward irradiance and evaluate their skills in order to estimate how well they perform compared to ~~in-situ~~ in situ measurements of PAR. The paper is organized as follows: in the Methods section, the Mediterranean Sea BGC-Argo floats network and the model configurations are presented. In the Results and Discussion section, we analyse the ~~1D~~ 1-D biogeochemical simulations and their sensitivity according to the objectives of the work. General remarks are illustrated in the Conclusions section.

2 Methods

2.1 BGC-Argo floats data

The Mediterranean Sea BGC-Argo array operating in the period 2012-2016 (Fig. 1) was composed of 31 floats that acquired 1314 vertical profiles ~~, Figure 1,~~ of temperature (T) and salinity (S), chlorophyll *a* concentration (Chl, units of $\mathrm{mg\,m^{-3}}$), derived from fluorescence measurements between 0 and 1000 m (see **??**), and radiometric quantities, such as downward planar irradiance (~~E$_d$~~ E_d) at three different wavelengths (λ = 380, 412 and 490 nm, units of $\mu\mathrm{W\,cm^{-2}\,nm^{-1}}$) and Photosynthetically Available Radiation (PAR, unit of $\mu\mathrm{mol\,quanta\,m^{-2}\,s^{-1}}$) integrated between 400 and 700 nm (**?**). Radiometric measurements were obtained in the upper 250 m, with vertical resolution of 1 m between 10 and 250 m and 0.20 m between 0 and 10 m. All profiles were acquired around local noon.

The quality control (QC) procedure of radiometric profiles was specifically designed to identify and remove the dark signal, atmospheric clouds and wave focusing at the surface (**?**). Note that the operational definition of PAR used by the BGC-Argo community takes into consideration the planar irradiance E_d rather than the scalar one E_o, therefore differing from its theoretical definition and leading to an underestimation of ~~its~~ E_o values by 30% or more (**?**). The scalar values of PAR were thus derived according to **?**, although the correction related to the irradiance scattering was neglected due to the lack of information on IOPs (see ~~section~~ Sect. 1 of ~~supplementary materials~~Supplementary material).

Vertical profiles of ~~chlorophyll~~ Chl concentration were quality-controlled according to the procedure of the international BGC-Argo program that removes spikes and corrects for non-zero deep values and non-photochemical quenching at the surface (**??**). Due to a factory calibration bias for WETLABS ECO series Chl fluorometers, Chl ~~a~~ concentrations were corrected by a factor of 0.5 (**????**).

All the data used in this study are freely available and compiled into the database published by **?**. To proceed with our study, 7 variables (T, S, Chl, ~~Ed380, Ed412, Ed490~~ $E_d(380)$, $E_d(412)$, $E_d(490)$, PAR) were vertically interpolated to a resolution of 1 m in the upper 400 m. Finally, we partitioned the profiles geographically into 13 (out of 16) subbasins (Fig1), with the majority of profiles located in the North Western Mediterranean (NWM, 332 profiles), followed by Northern Ionian (ION3, 172 profiles) and Southern Tyrrhenian (TYR2, 162 profiles). No data were available for the South-western Ionian (ION1) and the Eastern Levantine (LEV4) and only one profile was present in the Northern Adriatic (ADR1), as well as in the Western Levantine (LEV1). The WMO code specification for each BGC Argo float (along with their operational periods) is provided in the ~~section~~ Sect. 2 of ~~supplementary~~ Supplementary material.

2.2 1-D Biogeochemical Model

Biogeochemical processes have been simulated according to the voxel approach ("volume element with biological content and processes", **?**), discretized along the vertical in order to resolve vertical irradiance attenuation and nutrient gradients. Each voxel replicated light and mixing conditions according to the trajectory and measurements of the corresponding BGC-Argo float, thus simulating a pseudo-lagrangian experiment. No exchanges of mass between voxel and the surrounding field have been considered, which implies smaller mass exchanges due to horizontal diffusion and baroclininc components of the (upper

ocean) advection field compared to vertical processes and biogeochemical dynamics. Conversely, voxel exchanges heat with the atmosphere and receives light in accordance with its moving position. Such an approach, similar to the one adopted by **?**, has been already successfully applied by **?** in order to analyse BGC-Argo Floats in the North Atlantic.

Furthermore, it ~~was~~ is assumed that major biogeochemical transformations can be described by the Biogeochemical Flux Model (BFM) parametrizations (see below), properly driven by a bio-optical model, which has been validated by contrasting model results and experimental data, as shown later. The model is formulated through a system of partial differential equations:

$$\partial_t C_i(z,t) = \partial_z[D_v(z,t)\partial_z C_i(z,t)] + v_{sink,i}\partial_z C_i(z,t) + BFM_i(T,S,PAR,\overline{C}(z,t)) \quad (1)$$

where C_i is the i-th biogeochemical tracer simulated (i=1,50), D_v is the vertical eddy diffusivity derived with the vertical mixing model described in ~~subsection~~ Sect. 2.2.1, v_{sink} is the sinking velocity and BFM_i is the reaction term corresponding to the tracer C_i. T, S, PAR are the data measured by BGC-Argo floats. Since the surfacing of BGC-Argo floats is programmed at around local noon, the variability related to diurnal variation of solar irradiance is taken into consideration according to **?**.

The biogeochemical model BFM (**?**) is a biomass-based numerical model that simulates the biogeochemical fluxes of carbon, phosphorus, nitrogen, silicon, and oxygen, characterizing the lower trophic level (producers, consumers, and recyclers) of the marine ecosystem. Its application is based on the coupled transport-biogeochemical model OGSTM-BFM (**??**). It includes four phytoplankton functional types (diatoms, nanoflagellates, picophytoplankton, and dinoflagellates), carnivorous and omnivorous mesozooplankton, bacteria, heterotrophic nanoflagellates, and microzooplankton. Each variable is described in terms of internal carbon, phosphorus and nitrogen concentrations. Particulate and dissolved organic matter are also included, with the latter partitioned in labile, semi-labile and semi-refractory phases. The present study is focused mainly on Chl, reserving to future analysis (according to data availability and optical model complexity) a study of Plankton Functional Types (PFT) resource competition dynamics (**??**).

In ~~the~~ particular, the Mediterranean Sea Monitoring and Forecasting Centre (Med-MFC) operatively produces analyses, forecasts and reanalyses of a series of biogeochemical state variables (e.g. Chl, nutrients, pCO2) for Copernicus Marine Environment Monitoring Services (CMEMS) since 2015 using the MedBFM model (**???**), which assimilates surface Chl from satellite observations (**??**).

We tested a total of 17 classes of simulations (summarized in Table 1 and Table 2) :

- In the first set of simulations, the biogeochemical model was forced with PAR from BGC-Argo floats. Experimental values of temperature and density (computed from float profiles) were also taken into consideration. A simulation for each of the BGC-Argo float trajectories was performed with this set-up, hereafter abbreviated as REF.

- Four additional sets of simulations were performed on the REF configuration by applying different values of vertical eddy diffusivity coefficients (MLD1, MLD2, MLD3 and MLD4) in order to assess uncertainties due to different vertical diffusion parametrization.

– Six additional sets of simulations were performed by forcing the biogeochemical model with PAR obtained by alternative bio-optical parametrizations (OPT1, OPT2a,b,c,d), one of which (OPT3) considering also the current modeling approach in the ~~CMEMS Copernicus system (OPT3)~~Med-MFC. In this way, the possibility of using biogeochemical models in the absence of PAR measurements was assessed. ~~Finally, a~~

– A set of simulations was devoted to understand the impact of using a constant light approximation rather than following the diurnal light variation (CL1 and CL2 configurations) on ~~chlorophyll distribution.~~ Chl distribution.

– Furthermore, we evaluated the impact on light propagation due to coloured phytoplankton degradation products, i.e., CDOM, (OPT4a,b,c,d and OPT5). ~~We therefore tested a total of 17 classes of simulations that are summarized in Tab.1 and Tab.2.~~

~~The biogeochemical model BFM (?) is a biomass-based numerical model that simulates the biogeochemical fluxes of carbon, phosphorus, nitrogen, silicon, and oxygen, characterizing the lower trophic level (producers, consumers, and recyclers) of the marine ecosystem. Its application is based on the coupled transport-biogeochemical model OGSTM-BFM (??). It includes four phytoplankton functional types (diatoms, nanoflagellates, picophytoplankton, and dinoflagellates), carnivorous and omnivorous mesozooplankton, bacteria, heterotrophic nanoflagellates, and microzooplankton. Each variable is described in terms of internal carbon, phosphorus and nitrogen concentrations. Phytoplankton functional types can be characterized regarding prognostic Chl and can additionally consider the silicate component for diatoms. Particulate and dissolved organic matter are also included, with the latter partitioned in labile, semi-labile and semi-refractory phases. The present study is focused mainly on Chl, reserving to future analysis (according to data availability and optical model complexity) a study of Plankton Functional Types (PFT) resource competition dynamics (??).~~ Initial conditions for all biogeochemical variables of BFM are provided by the CMEMS reanalysis of Mediterranean Sea biogeochemistry (period 1999-2015, **?**) produced by the MedBFM model system. The initialization profiles for our 1-D configuration are extracted from the MedBFM model output array, taking the nearest model point to the BGC-Argo position in time and space.

Simulations' time scale corresponds to a typical BGC-Argo time-series length during the period 2012-2016, i.e. 11 months on average, with a vertical resolution of 1m. After being initialized, the model evolves without further assimilation of biogeochemical data from the 3D configuration.

Vertical eddy diffusivity coefficient profiles $D_v(z)$ are ~~here~~ represented as Gaussian-shaped functions ~~, using potential density values for the mixed layer depth (MLD)calculation with a density-based criterion (??). Such a shape is chosen due to its simplicity and in order to allow a~~ (see Sect. 2.2.1), thus allowing a gradual increase of vertical mixing through the pycnocline. Approaches and impacts of using different parametrizations to reconstruct mixing along the water column are shown and discussed in ~~section 2.2.1. Since the surfacing of BGC-Argo floats is programmed at around local noon, the variability related to diurnal variation of solar irradiance is taken into consideration according to ?.~~ Sect. 2.2.1.

~~Parameters derived for optical models using BGC Argo float data. For each version of OPT2 only data shallower than z_{max} were used to compute the regression.~~

Table 1. Model configurations considered in the present work. All simulations include diurnal variability except the two cases with continuous light (CL1 and CL2), which use 24-hour averaged irradiance values.

SIM	MODEL DESCRIPTION
REF	PAR from BGC-Argo floats ; $D_v^{background}$ = $10^{-4} m^2 s^{-1}$
CL1	as REF with continuous daily light
CL2	as REF with continuous daily light and $D_v^{background} = 10^{-6} m^2 s^{-1}$
MLD1	as REF with $D_v^{background} = 5\,10^{-5} m^2 s^{-1}$
MLD2	as REF with $D_v^{background} = 10^{-5} m^2 s^{-1}$
MLD3	as REF with $D_v^{background} = 5\,10^{-6} m^2 s^{-1}$
MLD4	as REF with $D_v^{background} = 10^{-6} m^2 s^{-1}$
OPT1	Riley: $K_d(PAR)$ = 0.04 + 0.054 $Chl^{\frac{2}{3}}$ + 0.0088 Chl
OPT2a OPT2b OPT2c OPT2d	4*$K_d(PAR)$ = a Chl^b + c
OPT3	$K_d(PAR)$ for the first optical depth z_{od} = z_{eu}/4.6
OPT4a	as OPT2a + Chl degradation to CDOM - time scale 1 day
OPT4b	as OPT2a + Chl degradation to CDOM - time scale 1 week
OPT4c	as OPT2a + Chl degradation to CDOM - time scale 1 month
OPT5	as OPT2a + CDOM following **?**

~~**Model** z_{max} R^2 **a b c** OPT2a 150 0.53 0.075 ± 0.0015 0.572 ± 0.018 0.027 ± 0.001 OPT2b 75 0.61 0.064 ± 0.0015 0.615 ± 0.021 0.040 ± 0.002 OPT2c 45 0.71 0.077 ± 0.002 0.469 ± 0.021 0.034 ± 0.002 OPT2d 30 0.75 0.088 ± 0.003 0.406 ± 0.023 0.029 ± 0.003~~

2.2.1 Vertical Mixing Models

~~Unlike radiometric data, vertical mixing is an indirectly obtained quantity, described in terms of~~ Vertical mixing is estimated from potential density (obtained from temperature and salinity data from floats) along the water column. Vertical eddy diffusivity coefficients (D_v) are defined as Gaussian-shaped functions in the form of:

$$D_v = D_v^{MLD} e^{-0.5(\frac{z}{(\sigma * MLD)})^2} + D_v^{background} \tag{2}$$

σ was identified after an initial tuning procedure and equals 0.3 in all simulations. Values in REF model are equal to D_v^{MLD} = ~~$1.0 m^{2s} \; 1$~~ $1.0 m^2 s^{-1}$ and $D_v^{background}$ = $10^{-4} m^2 s^{-1}$.

The mixed layer depth (MLD) was defined with the density criterion at the threshold value (**??**):

$$\Delta\rho_\theta = |\rho_\theta(10m) - \rho_\theta(z)| = 0.03 kg m^{-3} \tag{3}$$

In simulations MLD1, MLD2, MLD3, and MLD4, $D_v^{background}$ values were perturbed ~~for~~ by two orders of magnitude (from 10^{-6} to 10^{-4} m^2s^{-1}) in order to estimate the impact such variations have on modeled Chl profile shapes compared to measured ones (see Table 1).

2.2.2 Bio-Optical Models

Alternative parametrizations to measured PAR profiles were used in models OPT1, OPT2abcd, OPT3, OPT4abc and OPT5. They differ in methods used to evaluate the ~~Beer-Lambert~~ diffuse attenuation coefficient K_d(PAR), which is parametrized as a function of Chl concentration rather than being directly calculated from BCG-Argo irradiance data (see ~~Tab.1 and Tab.~~Table 1 and Table 2).

OPT1 uses the relationship obtained by a statistical analysis done by **??**:

$$K_d(PAR) = 0.04 + 0.0088\,[Chl] + 0.054\,[Chl]^{\frac{2}{3}} \tag{4}$$

In OPT2 models, statistical regressions were carried out between $K_d(PAR)$ and Chl measured by BGC-Argo floats at four different depth ranges: 150 m, 75 m, 45 m and 30 m (OPT2a to OPT2d, see Table 2 for details):

$$K_d(PAR) = a\,[Chl]^b + c \tag{5}$$

a and c represent regression coefficients and b the exponent (values reported in Table 2. Confidence intervals were calculated with Student's two-sided t-test, where the significance level α was set equal to 0.05). Diffuse attenuation coefficients $K_d(PAR)$ were calculated for PAR measured by BGC-Argo floats as the local slope of the natural logarithm of downwelling irradiance for layers of 15 m thickness for the euphotic depth range, which corresponds to an attenuation of downward planar irradiance to 1% of the subsurface value (**?**).

Albeit the regression ~~based on~~ in the upper 30 m ~~depth range measurements~~ of the water column showed highest correlation, all four bio-optical models were considered and adopted in simulations OPT2a,b,c and d (~~Tab.~~Table 2).

In model OPT3, based on the BGC-Argo data set, $K_d(PAR)$ is calculated for the first optical depth (**?**), the layer of interest for satellite remote sensing (**?**), and then adopted as a constant parameter for the entire water column. Such kind of light extinction definition has been used also in the ~~3-dimensional~~ 3-D version of the OGSTM-BFM model, which integrates $K_d(490)$ data from satellite sensors as the external optical forcing in the exponential formulation of downwelling irradiance ~~(for more details see **?**, section 2.2.3)~~(for more details see **?**, Sect. 2.2.3).

OPT4 and OPT5 models include CDOM dynamics as in the Mediterranean Sea the latter can absorb more than 50% of blue light (**??**), thus significantly impacting its attenuation along the water column. OPT4 assumes that CDOM is correlated to ~~chlorophyll~~ Chl production (**?**) and that the light attenuation is therefore affected by a progressive accumulation of ~~such a constituent~~ the latter ("dead" ~~chlorophyll~~Chl, initialized at zero concentration). In OPT4, accumulation is compensated by ~~decay (first order kinetic)~~ a linear decay set at different e-folding characteristic times: 1 day (OPT4a), 1 week (OPT4b) and 1 month (OPT4c).

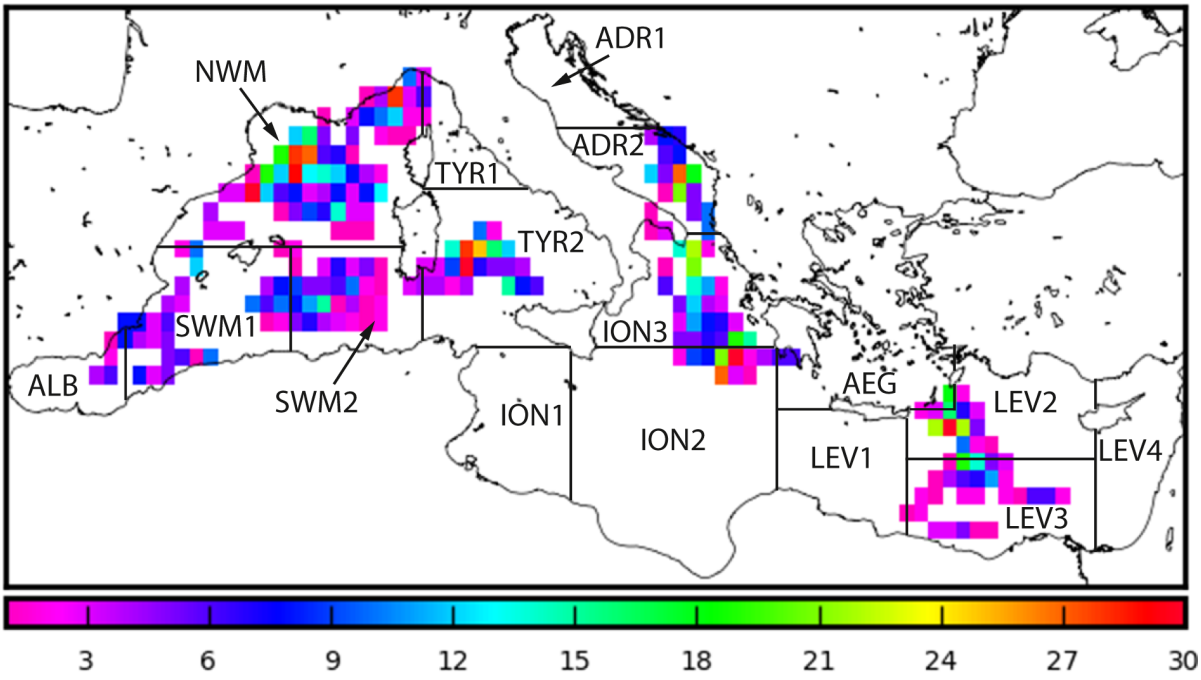

Figure 1. Spatial distribution of BGC-Argo float profiles superimposed to subbasin division used in the Mediterranean Copernicus Marine Environment Monitoring Service (CMEMS) system

Table 2. Parameters derived for optical models using BGC Argo float data. For each version of OPT2 the regression is performed in the depth range indicated by z_{max}. Data points are averaged for layers of 15 m thickness.

Model	z_{max}	R^2	**a**	**b**	**c**
OPT2a	150	0.53	0.075 ± 0.0015	0.572 ± 0.018	0.027 ± 0.001
OPT2b	75	0.61	0.064 ± 0.0015	0.615 ± 0.021	0.040 ± 0.002
OPT2c	45	0.71	0.077 ± 0.002	0.469 ± 0.021	0.034 ± 0.002
OPT2d	30	0.75	0.088 ± 0.003	0.406 ± 0.023	0.029 ± 0.003

OPT5 implemented a formulation of CDOM as described in **?**: a 2% fraction of all dissolved organic matter (DOM) fluxes is directed to CDOM, including both temperature-related decay and a photodegradation term based on PAR (**?**). ~~Additional investigations are provided in section 3.3 to discuss CDOM dynamics along the water column.~~ Given the mono-spectral nature of the current description of light, the attenuation of CDOM on PAR is computed by averaging the exponential law of CDOM absorption (**?**) on the visible range. Additional investigations are provided in Sect. 3.3 to discuss CDOM dynamics along the water column.

2.3 Statistical Analysis

According to the work's objectives, four classes of simulations were considered, which correspond to the following subsections: the reference simulation, a subset with perturbed vertical mixing models, tests with different optical configurations, and a last group of additional analyses involving CDOM description and diurnal variability. Outputs are validated qualitatively and quantitatively in terms of profile shapes and the deep chlorophyll maximum (DCM) depth. The DCM definition is based on the absolute maximum of Chl, excluding results of DCM shallower than 40 m or deeper than 200 m, as well as the ones with concentrations lower than 0.1 $\mathrm{mg\,m^{-3}}$. All results, both for model and BGC-Argo floats, are averaged on a weekly basis. Model outputs are compared ~~by means of~~ with a match-up ~~, Target~~ shown as target and Taylor diagrams (**?**). ~~In addition to the DCM depth, the performance in reproducing the DCM thickness and Chl concentration in the DCM layer were also analysed. DCM thickness is operationally defined through a Gaussian fit as from the maximum, thus the Chl concentration at DCM is averaged over the DCM thickness. For a couple of simulations (REF and CL1), skills are also compared at the surface layer (0 - 25 m). In order to avoid corrections due to non-photochemical quenching, profiles acquired only during stratified periods were considered. The target diagram~~ The former evaluates results with root mean square distance (RMSD) as the main statistical parameter, which was calculated following ~~equation~~ Eq. 6:

$$RMSD = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\left(m_i - o_i\right)^2} \tag{6}$$

where n is the number of data, m are the model ~~data~~ values and o are the observables.

Due to the various sources of possible uncertainties in the fluorescence-to-Chl conversion of BGC-Argo profiles (see Sect. 2.1 and references therein), we chose to focus our study on DCM depth rather than DCM magnitude. This is additionally justified with the BFM statistical sensitivity analyses (see Supplementary material), which considered DCM width, DCM magnitude and surface Chl. Results indicate that DCM depth is the most effective feature the model is able to reproduce.

3 Results and Discussion

3.1 Reference Simulation

The assimilation of PAR profiles ~~into the 1-D model~~ helped to accurately estimate the ~~deep chlorophyll maximum depth (Figure~~ DCM depth (Fig. 2). ~~The overall model skill in the REF configuration is shown in Figure 2, with the histogram within indicating~~

~~a normal distribution of residuals' deviation.~~ Measured and modelled DCM depth showed high correlation (r = ~~0.8~~0.83, p-value < ~~0.0005~~0.005). Both model and measurements indicate that DCM depth varies typically between 50-70 m in western areas (ALB, SWM1, SWM2, NWM, TYR) and is generally deeper in eastern areas (ADR2, ION3, LEV2, LEV3), between 100-140 m. Model tends to slightly underestimate the DCM depth variability (~~Figure~~ Fig. 2, regression slope = 0.81 < 1)~~, in fact, :~~ deepest simulated DCM are around 125 m depth, ~~whilst floats data~~ whereas data from floats reach 140 m (e.g. as ~~shown in Figure 2~~ lovbio018c~~data~~).

Chl patterns display high variability both at temporal and vertical scales ~~, shown in Figures~~ (Figs. 3 to 6). The subsurface Chl pattern is formed by patchy structures and ~~during stratification periods~~ it is generally deeper ~~moving eastward.~~ toward east during stratification periods (Fig. 6). BGC-Argo observations indicate that DCM is further eroded by vertical mixing occurring ~~generally~~ predominantly in autumn and early winter. At the ocean surface, the increase in Chl is triggered by rather shallow mixing ($\leq$ 75 m). Simulations provide an adequate reproduction of the Chl mixing timing and therefore of the DCM erosion. By comparing ~~point-to-point all~~ Hovmöller maps of all 31 floats (considering both depth and time variability) for measured and simulated Chl~~(examples are reported in Figures 3, 4, 5 and 6), ~~, a significant average correlation coefficient (r) of 0.75 is obtained: such a result quantitatively confirms that the alternation of mixing and stratification phases, as seen from BGC-Argo ~~chlorophyll~~ Chl measurements, is well reproduced. ~~At surface, the increase in Chl is triggered by rather shallow mixing (0-75 m layer). The initial condition statistics in reproducing DCM depth (R=0.62, slope of 0.53) is improved by the 1-D model for BGC-Argo data.~~

~~In addition to a correctly reproduced timing in the alternation of mixing and stratification phases, proved by high correlation, simulated~~ Simulated Chl reproduces also episodic signals, such as Chl deepening due to specific mixing events. For example, a mixing event in the NWM subbasin, reaching approximately 200 m depth during winter ~~in~~ 2015, triggers an intrusion of Chl (0.2 $\mathrm{mg\,m^{-3}}$) in deeper layers ~~consistently to~~ consistent with BGC-Argo float measurements (float lovbio067c, ~~Figure~~ Fig. 3). Similar dynamics is reproduced in winter 2014 (~~Figure~~ Fig. 4) for the lovbio035b float drifting from NWM toward the ALB subbasin.

Considering float trajectories, two kinds of situations are possible: the BGC-Argo float trajectory is relatively stationary in the deployment area (~~as shown in Figures~~ Figs. 3, 5 and 6), or the float ~~passively migrates extensively ,~~ migrates extensively by following a given water mass (~~as in Figure~~ Fig. 4). ~~It appears~~ Results confirm that also in the second case, when lateral ~~dynamics effects~~ advection processes could play an important role in ~~BGC-Argo float measurements, the approach applied~~ the float dynamics, the applied approach allows an adequate representation of measured Chl patterns. ~~However, it should be noted that in the~~ The present multi-float simulation~~there are no trajectories including~~ , however, does not include trajectories comprising both west and east Mediterranean basins~~. In such cases, ~~, where strong gradients between deep water nutrient inventories could invalidate the approach~~, thus~~ . In such cases, nudging or more sophisticated techniques would be required (**?**). ~~Lateral advection processes could indeed play an important role, although it appears that in the present case considering data-driven mixing and~~

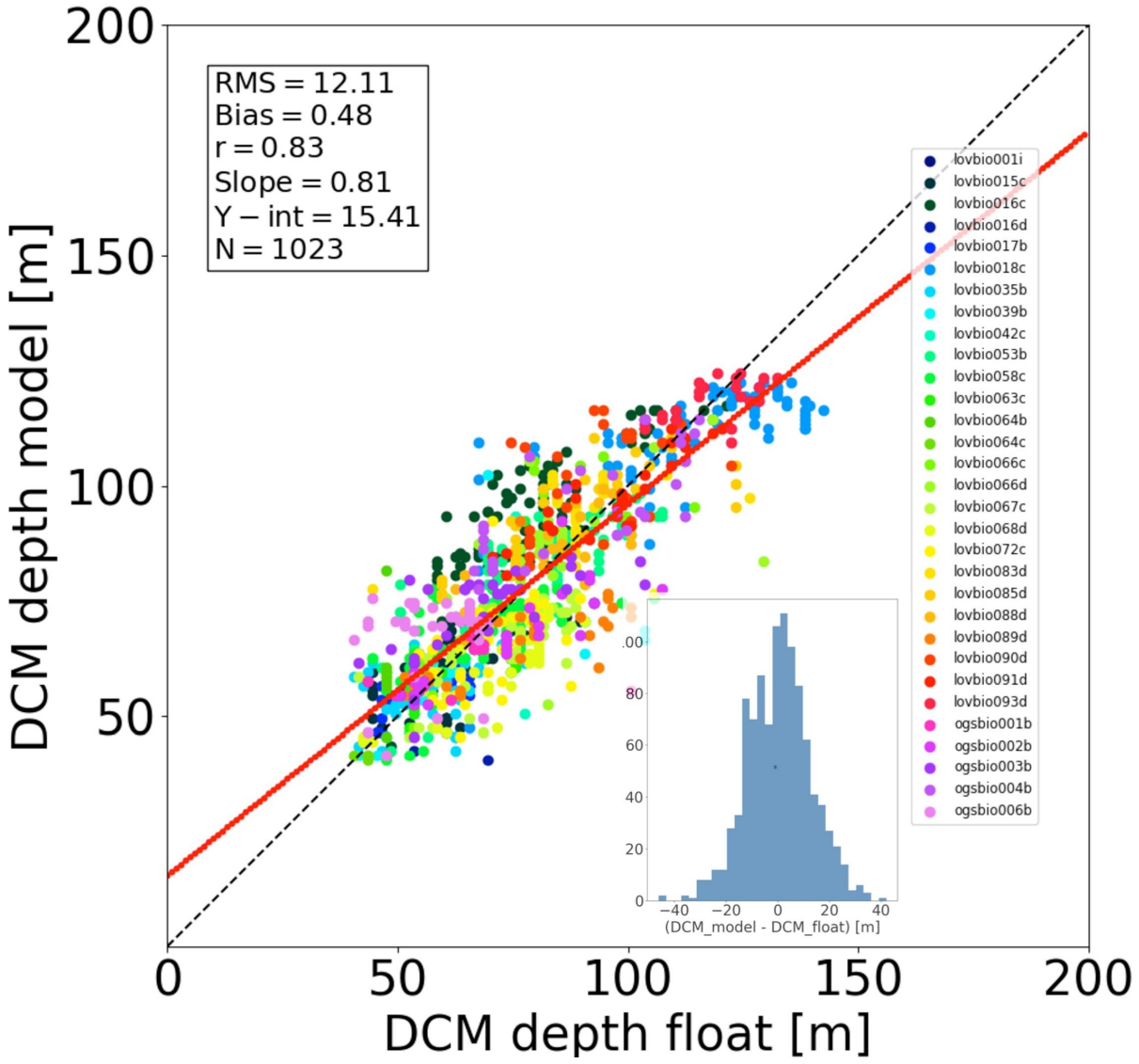

Figure 2. Match-up diagram comparing DCM depth obtained from BGC-Argo floats data versus REF model results. Each dot corresponds to a weekly profile. The red line depicts the linear regression between data and model values, defined by its slope and intercept (Y-int) shown in the box. Units of RMSD, Bias and Y-int are in meters. The correlation coefficient r is significant, with p-value < 0.005. The bottom sub-figure shows the residuals' histogram.

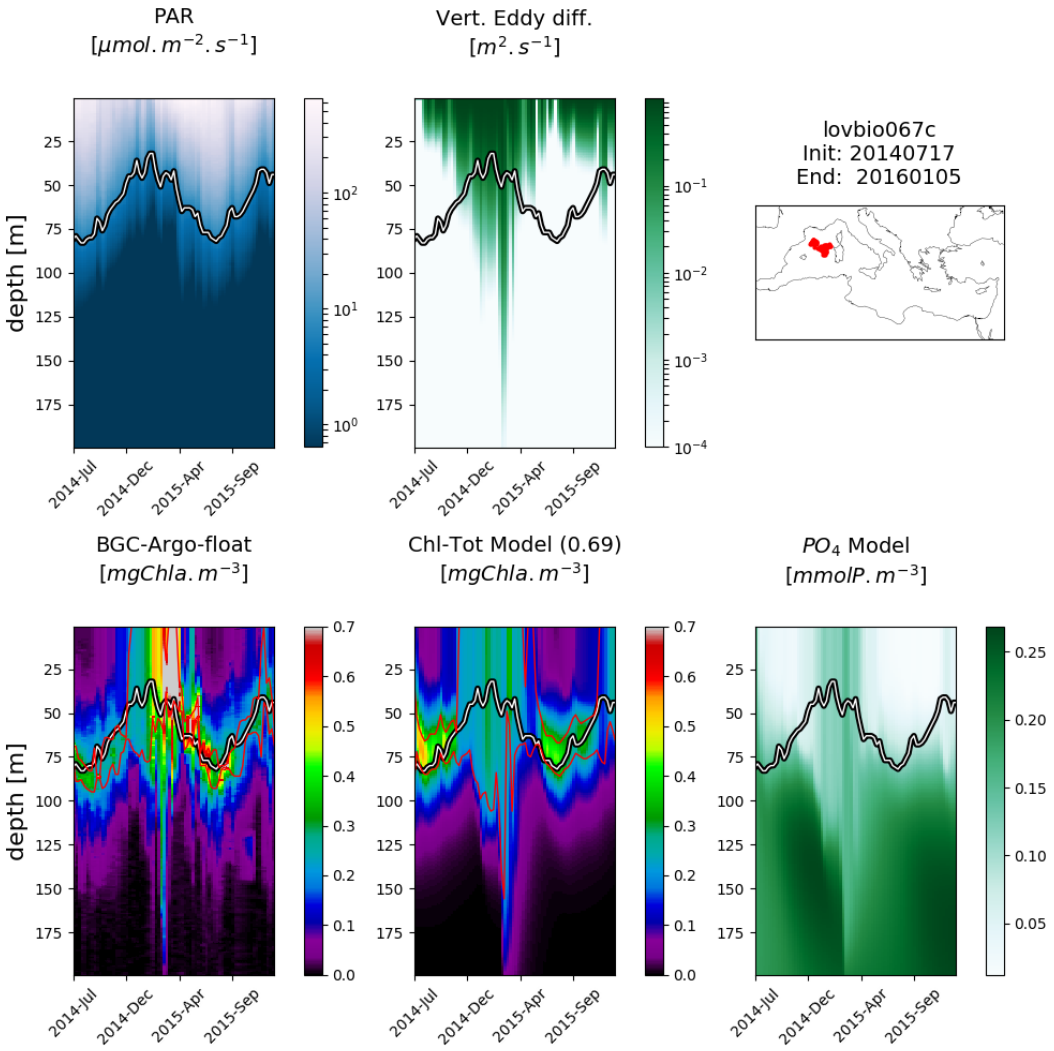

Figure 3. Hovmöller diagrams of BGC-Argo float lovbio067c (WMO code 6901649) comparing measured results and simulated ones (REF). The 6-imaged composite is organized as follows: top row shows PAR, vertical eddy diffusivity and the float trajectory; bottom row shows Chl derived from fluorescence measurements, simulated Chl and phosphate. The thick black-white line indicates the depth where PAR equals ~~0.5~~ 5.8 μmolquanta m^{-2} s^{-1} (**?**). The number in parentheses in modelled Chl indicates point-by-point correlation with BGC-Argo float Chl.

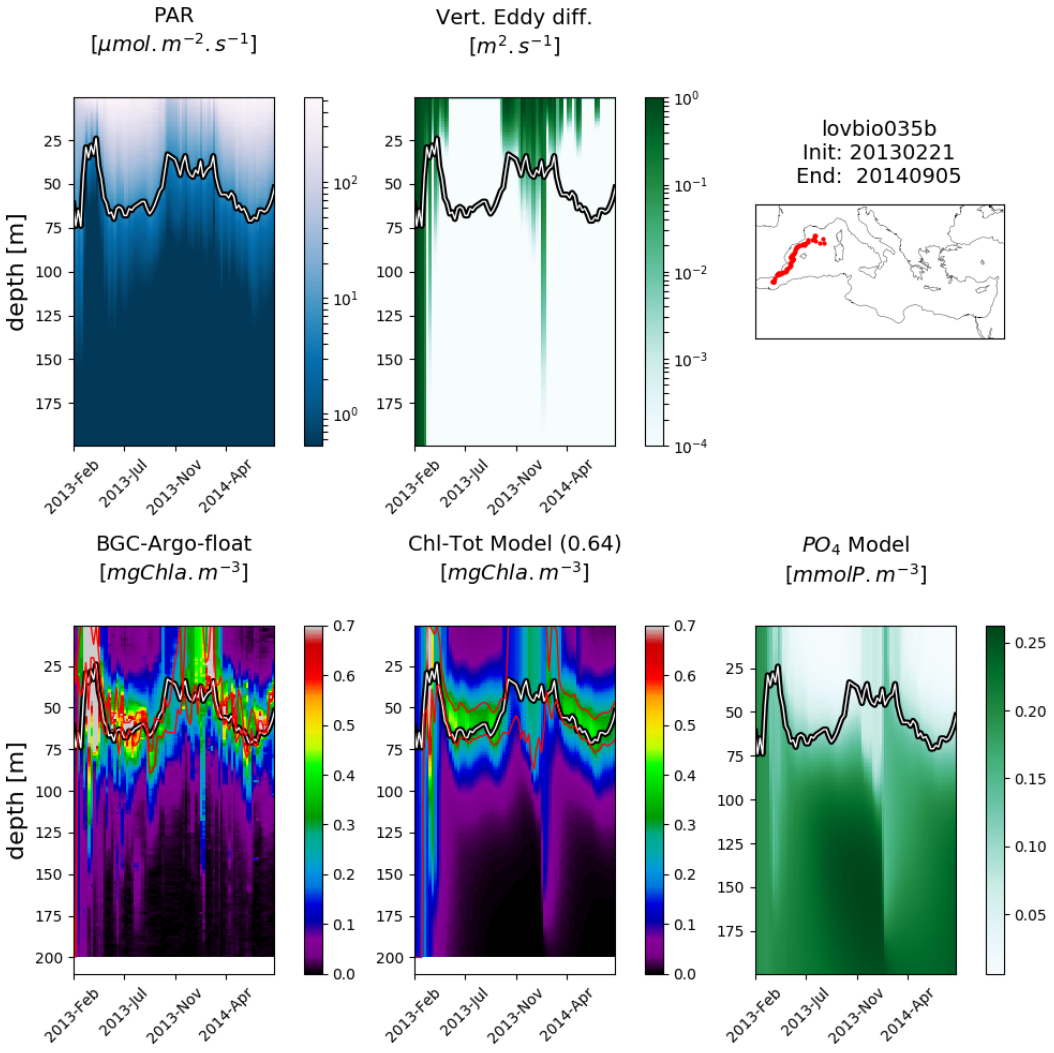

Figure 4. As ~~Figure~~ Fig. 3 but for the BGC-Argo float lovbio035b (WMO code 6901511).

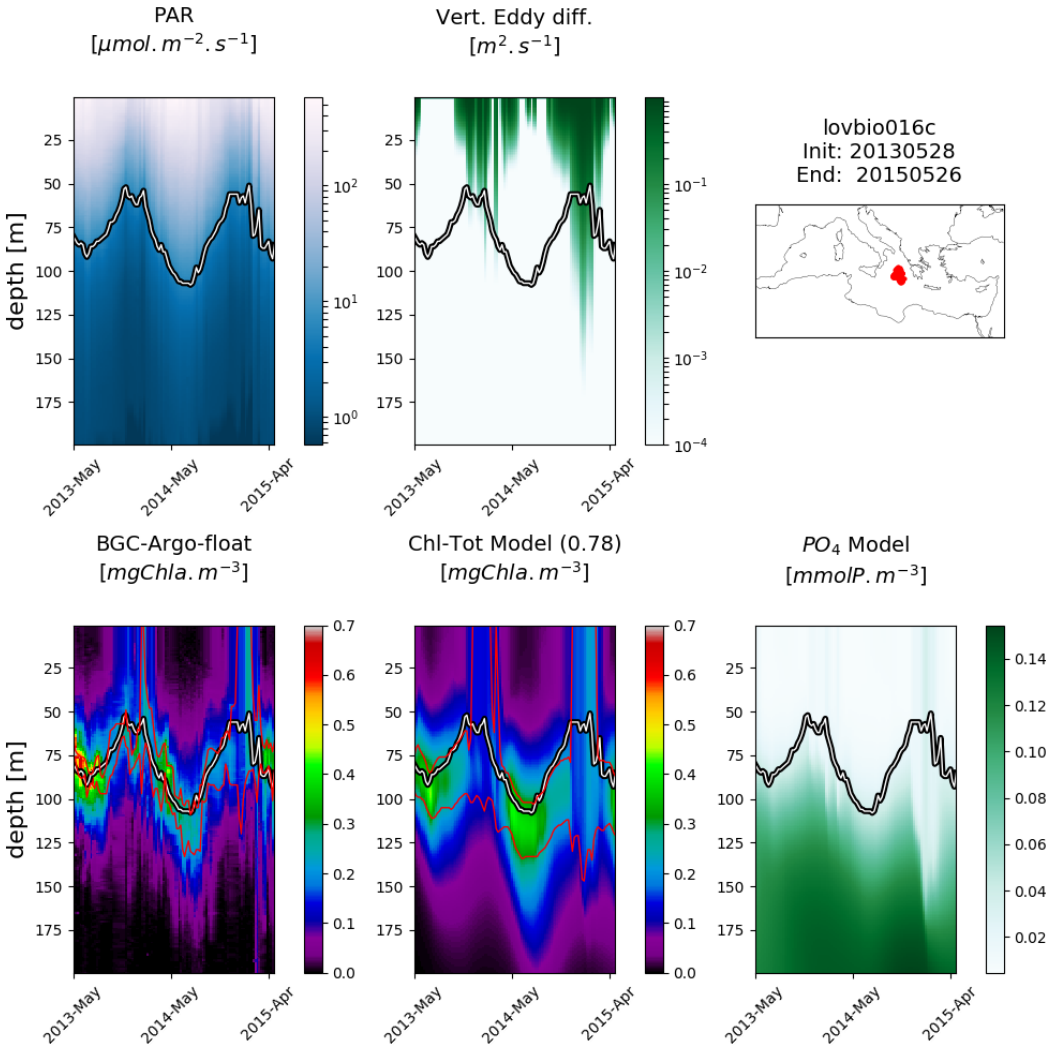

Figure 5. As ~~Figure~~ Fig. 3 but for BGC-Argo float lovbio016c (WMO code 6901510).

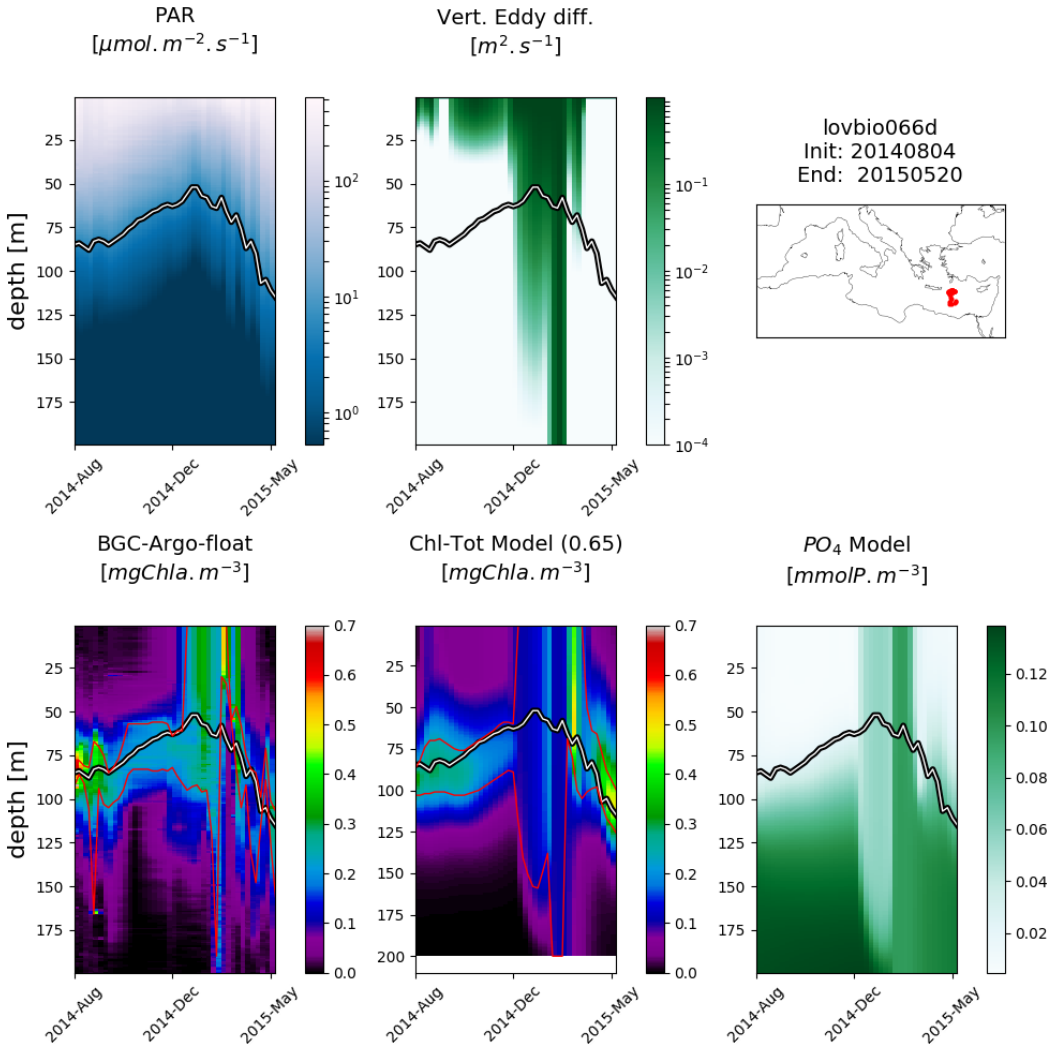

Figure 6. As ~~Figure~~ Fig. 3 but for BGC-Argo float lovbio066d (WMO code 6901655).

turbulence effects allow to simulate correctly the seasonal variability. The REF simulation can be therefore used as a reference for the following tests on mixing and bio-optical models analysed in the next sections. Furthermore, REF results

REF results further demonstrate that irradiance along the water column , besides mixing, is the driving mechanism controlling DCM depth . Figure 7a shows a besides mixing, which is proven by a significant correlation between DCM and euphotic depths (i.e. where irradiance reaches 1% of surface PAR), both in cases of both for measured and simulated Chl .

Panel a) DCM depth (z_{DCM}, x-axis) compared to the euphotic depth (z_{eu}, y-axis) both for modelled (red dot) and measured results (black dot). Red box (top left) reports statics for model z_{DCM} versus z_{eu}, whereas the black box (bottom right) shows statistics for z_{DCM} derived from chlorophyll data versus z_{eu}. Panel b) irradiance values (y-axis) at DCM depth (x-axis) both for modelled (red dot) and measured results (black dot). Horizontal blue line marks the 0.5 irradiance threshold (units) as identified in **?**.

Similar results, valid on annual average conditions, were found by **?** in their Eq. 9, where euphotic depth results to be 0.73% rather than 1% of surface PAR, interpreting (Fig. 7a). Such findings were already established in **?**, indicating that the DCM is located at a fixed PAR value, oscillating near the 0.5 isolume 5.8 $\mu\mathrm{molquanta\,m^{-2}\,s^{-1}}$ isolume (Fig. 7b, blue line). Comparable conclusions can be derived for analyses presented hereby. Data and model outputs show similar results with in the present study show a higher variability of critical PAR values in the case of shallower DCM , Figure 7. However, the model-predicted DCM seems strongly constrained by light regime, Figure **??**, whilst observed DCM fluctuates up and down over the euphotic depth (see transition from "T=29 weeks" to "T=31 weeks"). (Fig. 7b).

In order to further evaluate the dependence of model results on PAR forcing, two numerical experiments were carried out. In the first experiment, BGC-Argo floats were divided in couples composed by one trajectory located in the western basin and the other one in the eastern basin, by random selection. For each couple the initial conditions for nutrients were interchanged, which allows to estimate their impact on DCM depth. Results (see supplementary material, section 4)evidence that the inverted initialization of nutrients does not significantly alter the outcome in terms of DCM depth, resulting in a slope reduction from 0.81 to 0.62, and maintaining similar correlation and bias. Thus it appears that the role of nutrients is secondary compared to the impact of light on DCM depth regulation on such time scales.

Performing the same operation by switching light data instead of nutrients proves to be technically more complex, thus an alternative approach was applied, which consists of a sensitivity analysis similar to the one described in **?**. For this purpose, two BGC-Argo floats (lovbio018c and lovbio067c for east and west respectively) and phosphate and PAR parameters were selected, constructing an array of 21x21=441 simulations (per float) for bivariate perturbations. Such technique allows to further understand the driving mechanisms of DCM depth variability.

A perturbation of 50% of the of initial condition has only a minor effect on DCM depth position, Figure **??**, while changes in light conditions show a large effect (approximately 10 meters difference).

The same sensitivity analysis is used to evaluate model performance in reproducing DCM width and chlorophyll at DCM (plots for both lovbio018c and lovbio067c are included in the supplementary material, section 5). Results indicate that the DCM width has a variability of 6 m in the perturbation range (), as well as that the DCM magnitude is controlled by

nutrient availability rather than light. Comparing measured chlorophyll concentrations and model results shows that the skill in reproducing the DCM thickness is lower compared to the DCM depth (Figure **??**, r=0.55, slope=0.6).

Scatter plot of DCM thickness. Left panel reports REF simulation ($D_v^{background}$ =), right panel shows MLD04 simulation ($D_v^{background}$ =). The thickness is defined as centered on the maximum computed on the vertical profiles by means of a Gaussian fit. Scatter plot comparing 0-25 m average surface chlorophyll versus BGC-Argo float data for the stratified period condition (DCM > 40 m).

The DCM thickness varies between 20 and 40 m for the model, whereas a higher variability from 5 m to 40 m is measured by BGC-Argo (Figure **??**). In section 3.2 we evidence how DCM thickness is controlled by the background vertical eddy diffusivity coefficient ($D_v^{background}$). 0-25 m average surface chlorophyll shows similar skill (r=0.68, slope=0.63) as in the case of DCM thickness, Figure 10. The skill of the 1D model in reproducing DCM magnitude is lower than for the other indicators: measured chlorophyll concentration fluctuates in the DCM and the possible underlying mechanism (e.g. presence of Rossby waves or Kelvin waves) goes beyond the scope of the present paper. We show here the median chlorophyll in the DCM layer () for each BGC-Argo float (Figure **??**). In general, simulations tend to underestimate chlorophyll concentration compared to BGC-Argo float in the Western Mediterranean Sea . Following the procedure of the sensitivity analysis shown before, we evaluated the effects of perturbing nutrients for the BGC-Argo floats deployed in the West Mediterranean by increasing them by a factor 2 (orange dots, Figure **??**). The skill in reproducing the DCM depth is almost the same between REF and REF with higher nutrients (see supplementary materials section 6), however clearly showing the relevant impact on chlorophyll concentration at DCM. Therefore, it could be possible to finely tune the initial conditions to maximize both skills in terms of DCM value and DCM depth. However, considering the fact that the measurements of concentration of chlorophyll as derived from fluorescence present some uncertainties (**???**), we prefer to keep the initialization based on reanalysis.

Scatter plot of DCM chlorophyll concentration as defined in the text: median concentration of the REF (blue dots) and from the simulation increasing PO4 (orange dots).

The Mediterranean Sea is The Mediterranean Sea is a nutrient-limited basin (e.g. **???**), therefore an insight on the role played by nutrients requires further investigation. Phosphate dynamics shows an increase in surface Chl driven by nutrient uptake in upper layers due to convective mixing . (Figs. 3 to 6). During stratification periods, the phosphocline follows the euphotic layer threshold. From results shown hereby, it can be ascertained that together with a strong correlation between light and DCM depth, nutrient Nutrient concentration is an important driver in regulating phytoplankton biomass at DCM . Indeed, western subbasins exhibit significantly higher values, of both phosphate and biomass, compared to the eastern ones, Figure **??**. It should be noted that the Chl concentration at DCM (see sensitivity analysis in Supplementary material).

The REF simulation is forced by PAR measurements, hence we it is possible to evaluate the direct impact of nutrients' vertical fluxes (**?**) compared to light on DCM properties. The effect of self-shading by chlorophyll Chl and CDOM can increase the role of nutrients in terms of DCM depth modulation, which can be evaluated only by using bio-optical models where attenuation is regulated by chlorophyll Chl or CDOM as presented in section Sect. 3.3.

~~Example of a weekly time series of vertical profiles attributed to the REF simulation of lovbio035b BGC-Argo float (Figure 4) and compared to BGC-Argo float Chl values (thicker line). The horizontal dashed blue line represents the euphotic depth, whereas the horizontal dashed black line indicates the depth where measured PAR equals 0.5 as identified in ?.~~

~~Phosphate (x-axis) and total biomass concentration (y-axis) of phytoplankton at DCM depth, including all modeled float trajectories.~~

3.2 Vertical Mixing Models

As shown in the previous section, the vertical distribution of ~~chlorophyll~~ Chl displays a distinct variability, which can be at least partially ascribed to mixing. Typically, higher vertical eddy diffusivity values imply smoother structures. During the stratification phase, when DCM forms, the controlling mixing parameter is the background diffusivity $D_v^{background}$.

Simplified theoretical models, such as the KiSS ~~(after the names of ??)~~(??), can provide rough quantitative scales in order to determine the minimum vertical length scales (L_0) that allow the formation of stable biomass patches (?), including the DCM, in a steady state hypothesis:

$$L_0 \propto \sqrt{\frac{D_v}{\mu}} \tag{7}$$

where D_v is the vertical diffusivity coefficient and μ is the growth rate ~~;~~ (in stratified conditions ~~,~~ $D_v = D_v^{background}$~~.~~ ~~Considering any compact vertical interval with favourable conditions for plankton growth (in terms of irradiance and nutrient availability), the~~). An increase of background diffusion over a critical value will produce a dispersal of patchy structures (i.e. a relative maximum of ~~chlorophyll~~ Chl concentration), whereas an increase in growth rate μ can drive the formation of finer scale structures~~by a reduction of L_0~~.

The dynamics presented in this study is~~much more complex compared to~~ , however, more complex than KiSS, both in terms of BGC-Argo ~~floats data and in~~ data and the 1-D ~~medium-complexity biogeochemical model (BFM)~~BFM model. Vertical eddy diffusivity can simultaneously affect nutrients, phytoplankton, and mesozooplankton with intricate interactions, which in turn make difficult to derive analytical solutions. Moreover, unlike KiSS, both the model and environment are hardly ever in a steady state condition, as a result of daily and seasonal oscillations in physical forcings, which are essentially due to variability in diel irradiance and vertical mixing.

Several simulations, labelled as MLD1, MLD2, MLD3 and MLD4, were carried out by changing the background vertical eddy diffusivity coefficient ~~$D_v^{background}$ values~~ values ($D_v^{background}$) by two orders of magnitude~~(from~~ , i.e. from 10^{-6} to $10^{-4}\,\mathrm{m^2\,s^{-1}}$ ~~, see~~ (Table 1). This subset of simulations (with float-derived PAR) clusters at a correlation of approximately 0.8 with ~~a root mean square difference (RMSD)~~ RMSD of DCM depth between 10-15 m. ~~Modeled chlorophyll profiles appear much smoother than the observed ones, following a Gaussian shape for all tested values of eddy diffusivity. Small scale patterns are not detectable even when~~ Perturbing $D_v^{background}$ ~~values are reduced to a minimum. Further analyses concerning these aspects are shown in section 3.4.~~ over two orders of magnitude (from REF to MLD4) shows that the impact on DCM

position is lower than 10 m (Fig. 8b), with an uplift of DCM depth with higher $D_v^{background}$ (**?**). The direct impact of eddy diffusivity appears lower compared to direct light modulation on DCM depth.

3.3 Bio-Optical Models

The alternative bio-optical models (OPT1, OPT2, OPT3) were slightly less accurate compared to REF~~: correlation decreases~~, with correlation decreasing from 0.8 to 0.6-0.5 ~~, Figure 8~~(Fig. 8a). The OPT3 simulation showed a bias very close to zero, thus suggesting an intermediate skill compared to assimilated PAR simulations (e.g. REF) and the bio-optical models (OPT1 and OPT2). OPT1 and ~~the~~ OPT2 cluster of simulations show slightly lower correlations with a RMSD of approximately 20 m in all cases, with an increase in bias (almost zero for OPT1 and from 6 m ~~(OPT2a)~~ to -14 m ~~(OPT2d)). The latter may stem from the fact that statistics performed~~ for OPT2a to OPT2d~~models ranged from 150 m to 30 m respectively, therefore lowering the number of data considered due to a reduced depth interval. Despite an increasing correlation of the bio-optical model linear regression with decreasing depth range, it should be underlined that the equations for lower depth ranges (such as OPT2d for the first 30 m)most likely do not perform well at greater depths, hence a higher bias in spite of a higher correlation coefficient~~). ~~The ensemble of simulations with alternative optical models shows in all cases smoother curves compared to measured Chl profiles (see Figure **??**). Chlorophyll self-shading effect increases from OPT2a to OPT2d, as explained above, due to different depth ranges of the dataset used to compute linear regressions.~~

Some of the bio-optical models considered, in particular OPT1, OPT2a and OPT2b, reproduce the DCM depth gradient between western and eastern subbasins with a tolerance of ± 10 m (~~Figure~~ Fig. 9). In previous studies (**??**), the correct simulation of the DCM depth longitudinal gradient was obtained by forcing the system with a space-time dependent light attenuation parameter based on Secchi disk climatology or on satellite $K_d(490)$ data. Both empirical approaches prevent to understand whether the origin of such gradients is directly related to external forcings or ~~, on the contrary,~~ if it can be interpreted as a self-emerging property ~~,~~ (i.e. related to the appearance of features which are not directly and explicitly imposed from the choice of boundary conditions or model parameters used in the numerical experiment (**?**)~~.Results shown in Figure 9a~~). Results suggest that a gradient in DCM depth could be partially reproduced and explained in terms of internal biogeochemical processes and partially due to external forcings (i.e. downward irradiance and nutrient initial conditions), even without considering lateral dynamics ~~.~~(Fig. 9a).

A direct analysis of the impact of alternative bio-optical models on light attenuation ~~, Figure 9b,~~ (Fig. 9b) indicates that the simulated eastern basin waters present generally lower K_d values (and lower dispersion around the median) for REF and OPT3. In other cases, where self shading is included, the variability is driven by ~~chlorophyll~~ (Chl (from OPT1 ~~, OPT2a, OPT2b, OPT2c, OPT2d and OPT4a, OPT4b,~~ to OPT4c) or by ~~chlorophyll~~ Chl and CDOM (OPT5), as bio-optical model parameters do not depend on space and time explicitly. West-east gradients are higher for maximum light attenuation along the water column (cross mark, ~~Figure~~ Fig. 9b) where ~~the concentration of chlorophyll~~ Chl concentration is higher. ~~Note that for the~~ For OPT3, the average and maximum K_d overlap since K_d is ~~for this simulation~~ parametrized as constant along the water column.

~~In fact, the~~ The average surface PAR of the ~~dataset~~ data set considered is higher in ~~the~~ eastern areas, especially during the months of January (40%), September (15%), October (22%), November (36%), December (16%), probably due to clearer at-

mospheric ~~weather~~ conditions. During summer, when DCM stabilizes, the west-east differences in measured surface PAR are lower and oscillate around 10%, however still contributing in increasing irradiance penetration at deeper layers.

The western and eastern ~~basins~~ subbasins are also different in terms of nutrient regimes that in turn impact biogeochemical dynamics and the DCM depth gradient in non-trivial ways. ~~In particular, the~~ The role of nutrients can be evaluated by perturbing initial conditions for the trajectories starting in the western subbasin, as shown in ~~section 3.1~~the sensitivity analysis reported in the Supplementary material. Results indicate that increased nutrients in the western subbasin cause an amplification of the west-east light attenuation gradients (~~Figure 9c~~Fig. 9d) related to the increase of ~~chlorophyll. The OPT2a test (with increased nutrients) appears to be the most consistent one compared to REF and OPT3, in terms of Kd west-east gradients.The~~ Chl.

The emerging conceptual scheme is that the first-order controlling mechanism for DCM depth is related to light propagation along the water column, as shown in REF and OPT3 simulations. Other tests indicate that nutrients modulate K_d consistently with gradients simulated in REF. The decadal temporal scale of subsurface nutrient variability (**?**) controlling self-shading mechanisms is longer than the one of simulations, suggesting that ~~such a mechanism~~ the role of nutrients in DCM positioning is especially regulated through initial conditions chosen for the present simulations.

Another key factor pertains to shorter wavelengths (400-450 nm) in the visible part of the spectrum: when light penetrates deeper along the water column, compounds like CDOM are more effective in absorbing light and might in turn enhance spatial gradients in irradiance regimes, which could synergistically contribute to a deeper DCM in eastern subbasins. However, with a current monospectral formulation, such aspects still cannot be addressed. ~~Multi-spectral~~ Multispectral configurations linked with specific PFT and CDOM absorption terms are thus needed for future in-depth studies of the questions raised in the present work (**?**).

3.4 Daily variable versus constant PAR forcings

The use of daily averaged irradiance (i.e. with continuous light, CL1 and CL2) was compared against REF that includes the diurnal variability. A consistent reduction of surface Chl concentrations was observed in the former case (~~Figure~~ Fig. 10), with a correlation lower than REF, affecting much less (in relative terms) the values around DCM (~~CL2 is shown in Figure~~ Fig. 11).

Near the surface, phytoplankton is limited by low nutrients (especially in eastern subbasins) whereas closer to DCM, the trophic limitation is weaker, sometimes ~~null~~ nonexistent (**??**). One possible explanation could be that light limitation at ~~DCM at a~~ the DCM at low-irradiance ~~regime~~ values is almost linear, thus the ~~averaging effects appear to be having a smaller impact than at~~ PAR daily averaging effects have a larger impact at the surface, where light limitation is highly non-linear due to saturation. Furthermore, the ~~Geider~~ BFM formulation for Chl acclimation (**?**) in case of diurnal variability generates an increase in ~~chlorophyll-to-carbon~~ Chl-to-carbon (Chl:C) ratio. This could in turn have important consequences in operational applications, where data assimilation is employed for model skill improvement~~. At~~ : at the surface, the adoption of a diurnal cycle formulation could reduce corrections made by the assimilation scheme and therefore minimize possible spurious trends introduced by it (**?**).

Combining daily-averaged irradiances with lowest diffusivity rates ($D_v^{background}$=$10^{-6}\,m^2\,s^{-1}$, simulation CL2) results in additional relative ~~chlorophyll~~ Chl maxima at surface layers (~~see Figure~~ Fig. 11, panel "T = 33 weeks"), as well as in increased patchiness of the ~~entire~~ whole vertical profile. Similar ~~chlorophyll~~ Chl profiles with multiple subsurface maxima were identified in a comprehensive fluorescence data analysis in the Mediterranean Sea ~~, ?. Theoretical consideration predicts~~ (?). Theoretical considerations predict different maxima along the water column based on the Tilman resource competition theory applied to a heterogeneous system (**?**). At this stage, however, it is difficult to assess whether the patchy structures observed in data and model are, for various reasons, realistic or ~~artefactual~~artifactual. Nonetheless, it can be ascertained that the background diffusion needed to maintain such structures in model simulations is very low.

~~Within~~ As a result, within the framework of currently used mathematical formulations in the 1-D BFM model, the inclusion of diurnal variability tends to reduce the formation of fine-scaled structures that could be interpreted in terms of a reduction in diel growth (μ) or seen as a possible perturbation that has an equivalent effect of an increased diffusion.

3.5 Bio optical models with CDOM formulation

OPT4 and OPT5 simulations take into consideration CDOM dynamics by including an additional term in OPT2a, where light attenuation by PAR was described only in terms of Chl. In OPT4a, b, and c, CDOM is parametrized as "dead" ~~chlorophyll~~Chl, by changing only the rate of Chl decay from 1 day to 1 month. Such simplified dynamics description ~~, albeit arguably, derives from~~ derives from the high correlation observed between Chl and CDOM in **?**. ~~It should be noted, however, that no analysis , which could corroborate findings from ?,~~ However, no analysis was carried out within the ~~dataset examined hereby~~ present data set to corroborate findings from **?** due to a lack of information on CDOM fluorescence. In all three model configurations, the "dead" ~~chlorophyll~~ Chl accumulation results in higher turbidity levels that in turn reduce light penetration depths. This is quantified by significantly negative DCM biases (over 40 m in OPT04c), which result in shallower DCM compared to BGC-Argo derived profiles since the attenuation of Chl is overestimated even when considering fastest degradation rates. (~~Figure~~ Fig. 8). The experiment OPT5 mimics the CDOM dynamics described in **?** where a lower bias is observed compared to the (over)simplified OPT4 tests (where correlation coefficients range from 0.6 to less than 0.1 for OPT4a to OPT4c respectively). OPT5 still results in a negative bias of around 10 m compared to the values from -25 m to -40 m for OPT4a to OPT4c.

~~In open ocean systems, at least three different mechanisms concerning CDOM entrainment in the euphotic layer are considered: lateral flux of CDOM from terrestrial waters (allochthonous origin), production of CDOM within the euphotic layer (autochthonous origin) and bottom-up flux of CDOM from the subsurface layer not affected by bleaching (?).~~

~~Figure 12 shows an example for a BGC-Argo float deployed in the North West Mediterranean subbasin (NWM).~~ The model, regardless of initial conditions, correctly drives CDOM absorption coefficients in deeper layers to low values, while an enhanced surface production reinforces mineralization and bleaching ~~and thus realizes a continuum of CDOM reactivity and lability.~~ (Fig. 12). Results of CDOM variability from the BOUSSOLE site (north-west Mediterranean, **?**) show that CDOM absorption ranges to a maximum value of 0.07 m^{-1} and indicate that there is a temporal delay between phytoplankton bloom and a maximum in CDOM absorption (~~Figure~~ Fig. 3 in **?**), whereas deeper layers (below 100 m) have generally lower CDOM absorption. The ~~dataset~~ data set shown in **?** evidences that cycles of CDOM accumulation are followed by depletion in the

upper 10 m due to photodegradation in summer. In ~~modeling results presented hereby~~the model results presented here, bleaching has a deeper effect over the entire CDOM "productive" layer (see red and blue lines, ~~Figure~~ Fig. 12), while the subsurface CDOM maximum is not reproduced. ~~Additional investigations of the OPT5 model configuration can address the autochthonous source dynamics, as well as the bottom-up flux of CDOM in this region.~~ The lack of CDOM accumulation in deeper layers for the OPT5 configuration hinders a proper analysis of the mechanism suggested in ~~section~~ Sect. 3.1 related to the emergence of CDOM from subsurface dark layers. Improving model dynamics calibrations could be possibly achieved by utilizing information on CDOM light absorption from BCG-Argo floats measurements (**??**).

4 Conclusions

The coupled modeling/experimental approach presented here provides a robust and accurate reproduction of the DCM variability across the Mediterranean Sea. Such a combined configuration can integrate in a single framework multi-data measurements provided by BGC-Argo floats. DCM is a ubiquitous feature of the Chl vertical structure in the Mediterranean, and different forcing conditions generate geographical gradients in DCM characteristics (i.e. shallower DCM in western regions, deepening eastwards). Second-order features, such as impulsive vertical spikes or specific patterns observed in BGC-Argo profiles, are also qualitatively reproduced. Results for the reference simulation, where measured PAR is adopted, are summarized as follows:

- mixing and irradiance propagation control Chl dynamics;
- DCM position is mostly controlled by PAR.
- nutrients control the amount of biomass at DCM.

~~Moreover, it~~ It was demonstrated that vertical processes considered in the 1-D model, such as irradiance regimes and vertical mixing, allow to properly reconstruct a large part of Chl dynamics, which was quantified also by skill diagrams.

~~The~~ Moreover, the role of nutrients in modulating self-shading (as inferred with bio-optical alternative experiments) appears relevant to shape west-east heterogeneity of vertical light attenuation.

The emerging conceptual scheme is that DCM gradients are directly controlled by irradiance modulation, ~~than is~~ in turn controlled through bio-optical processes which change attenuation according to optically active substances (e.g. ~~chlorophyll~~Chl, CDOM). Nutrients can impact attenuation by regulating ~~chlorophyll content~~Chl concentrations. The time scale of the ~~nutrient pool~~ subsurface nutrient inventory variability is longer than the ones considered in the present ~~simulation, thus enabling initial conditions to modulate~~ simulations, therefore initial conditions have an impact on west-east gradients.

Such kind of data-rich experiments, combined with a 1-D numerical model, could be considered as a useful tool also to a broader community, rather than only to biogeochemical modelers, in particular to address process studies.

The presented approach might be ~~useful also~~ also strategical to quantify the amount of measured signal related to vertical dynamics and the one derived from other processes, such as horizontal advection and subduction of water masses. The usage of PAR measured from BGC-Argo floats (used in REF, CL1, CL2, MLD1, MLD2, MLD3 and MLD4) provides higher correlations

compared to configurations with alternative bio-optical models (used in OPT1, OPT2, OPT3, OPT4 and OPT5). CL1 (without diurnal cycle) shows overall highest correlation, comparable with REF~~(Figure 8a).~~

. The comparison of different bio-optical models indicates that, when lacking direct measurements of PAR in subsurface layers, the most fitting alternatives would be OPT3, OPT2a and OPT1, ~~that provide relatively~~ resulting in lower bias and higher correlation coefficients (between 0.5 and 0.7), as well as lower RMSD values compared to REF.

~~Such an~~ Our analysis can also ~~suggest the rate of improvement when considering a value of~~ help determine how the use of light fully integrated in the visible range of the spectrum (400 to 700 nm, REF) ~~versus~~ improves predictions when compared to simplified approaches (i.e. all the OPT simulations here considered).

These results ~~further support~~ also highlight the strategic relevance of BGC-Argo data~~. Temperature~~: temperature, salinity and radiometric parameters encapsulate fundamental information for the reconstruction of primary producers dynamics and are paramount to investigate hypotheses concerning DCM formation. CDOM fluorescence data measured by BGC-Argo floats could be integrated in simulations to further infer and reconstruct the observed biogeochemical processes.

~~Furthermore, considering~~ Considering a general 3-D biogeochemical model, it is not possible to have a full data coverage of the in-water PAR field without a fully coupled radiative transfer model. Such an approach could be thus exported to more complex ~~3-D~~ biogeochemical models and generalized at a global scale.

Code and data availability. The BFM biogeochemical model and its documentation can be downloaded at the following address: http://bfm-community.eu/. The quality-controlled databases used in the present manuscript are publicly available from the SEANOE (SEA scieNtific Open data Edition) publisher at https://doi.org/10.17882/49388 and https://doi.org/10.17882/47142 for vertical profiles and products within the first optical depth, respectively.

Author contributions. E.T. and P.L. have designed the manuscript. E.T. performed the BGC-Argo data analysis, P.L. performed the simulations. All the authors contributed to the manuscript writing.

Competing interests. Authors declare that no competing interests are present.

Acknowledgements. This work is part of the PhD project of Elena Terzić, which was funded under the CMEMS contract for the Biogeochemistry Production Unit for the Mediterranean Sea. The simulations were performed in the framework of the ISCRA C project NOVBIOGE (HP10C8C9O6), granted by CINECA, Italy. This work was supported by the French "Equipement d'avenir" NAOS project (Novel Argo Ocean observing System - NAOS project funded by Agence Nationale de la Recherche (grant agreement ANR J11R107-F); the "Remotely-sensed biogeochemical cycles of the oceans- remOcean" project funded by the European Research Council (grant agreement 246777); the Argo-Italy project funded by the Italian Ministry of Education, University and Research; and the French Bio-Argo program - Bio-

Argo France funded by CNES-TOSCA, LEFE Cyber, and GMMC. We acknowledge sponsorship from the ~~BIOPTIMOD CMEMS Service Evolution projectand from the~~ MISTRALS-MERMEX project. This work has been carried out as part of the Copernicus Marine Environment Monitoring Service (CMEMS) BIOPTIMOD project. CMEMS is implemented by Mercator Ocean International in the framework of a delegation agreement with the European Union.

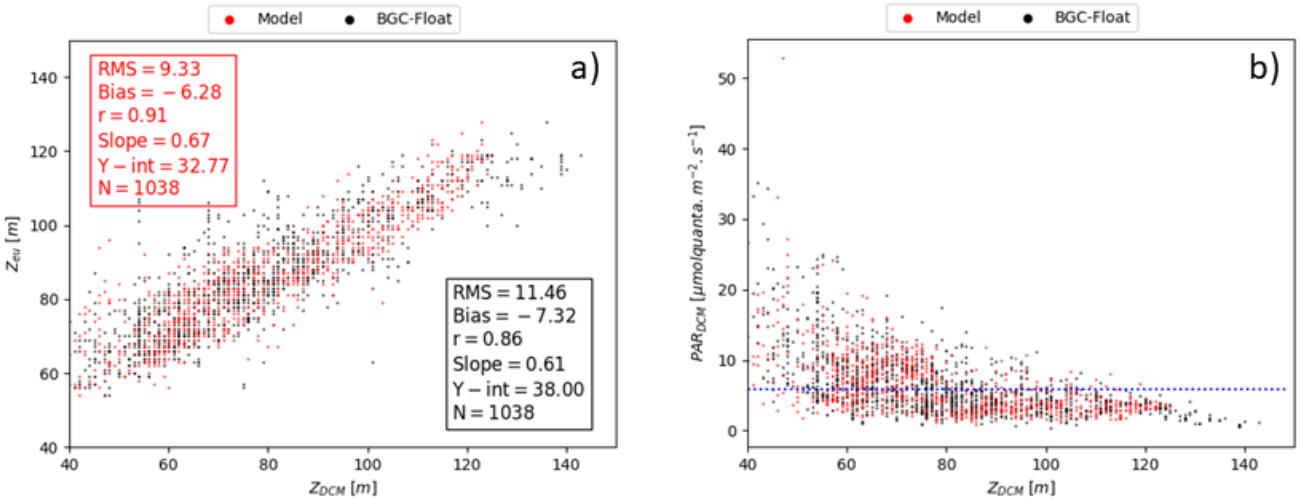

Figure 7. ~~Sensitivity analysis of~~ Panel a): DCM depth ~~perturbing light and initial conditions of PO4 (both by an uniform factor reported on axis in percentage~~ z_{DCM}, x-axis) ~~along~~ compared to the ~~water column~~euphotic depth (z_{eu}, y-axis) both for modelled (red dot) and measured results (black dot). ~~'R' marks~~ Red box (top left) reports statistics for model z_{DCM} versus z_{eu}, whereas the ~~reference values~~black box (bottom right) shows statistics for z_{DCM} derived from Chl data versus z_{eu}. ~~The BGC-Argo float here reported is the lovbio018c. Each pixel is a full simulation of a total of 21x21 simulations. The~~ Panel b): irradiance values (y-axis) at DCM depth ~~is averaged over~~ (x-axis) both for modelled (red dot) and measured results (black dot). Horizontal blue line marks the ~~simulation period~~5.8 irradiance threshold (units $\mu\mathrm{molquanta\,m^{-2}\,s^{-1}}$) as identified in **?**.

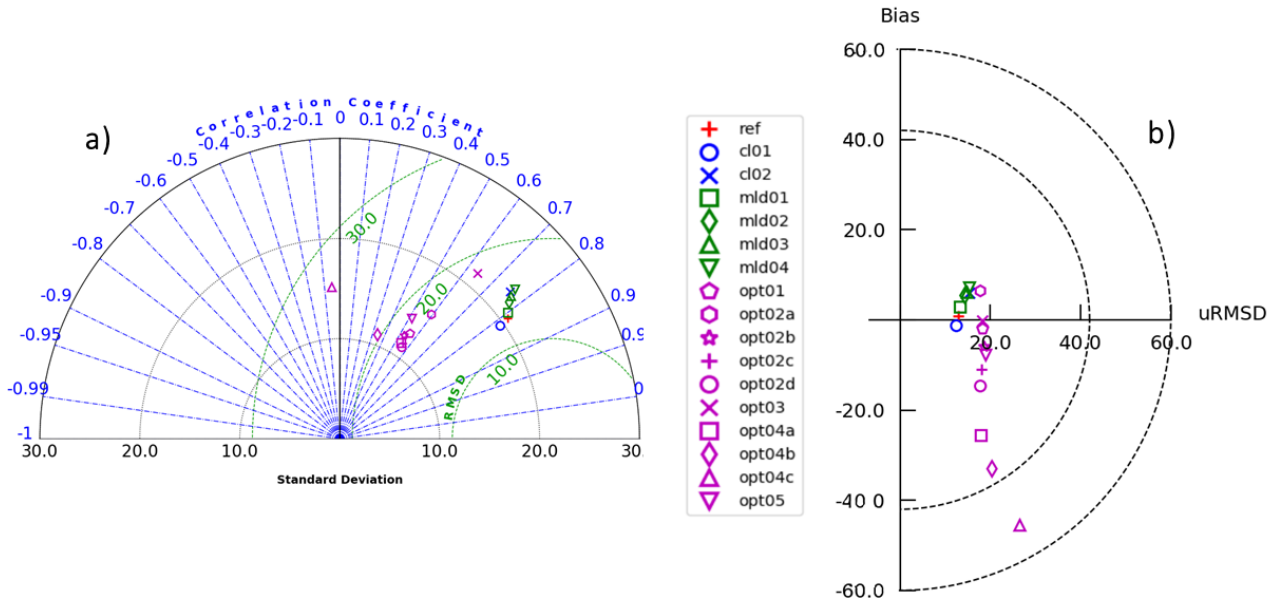

Figure 8. Panel a): Taylor diagram showing model skill in reproducing DCM depth compared to data. Correlation is represented by the angle with positive x axis, whereas distances from the origin depict standard deviations. Green circles illustrate iso-contours of RMSD levels. Panel b): Target diagram showing model skill in reproducing DCM depth compared to data. Distance to the origin defines the RMSD, all units are in meters. The position on x-axis is positive if the model standard deviation is higher than the one from data results and negative in the opposite situation. For the sake of completeness, all models considered are reported in these summarizing skill diagrams.

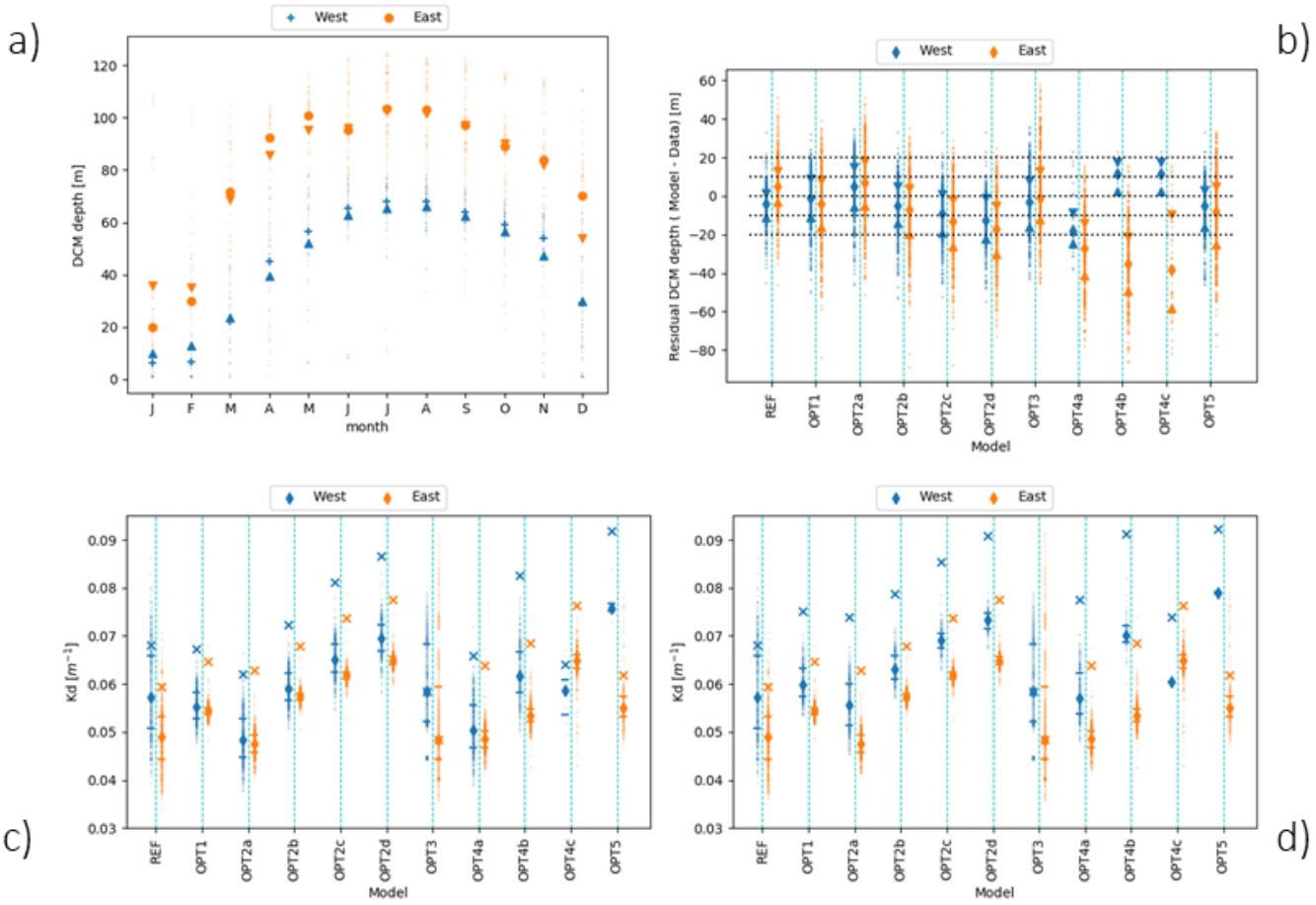

Figure 9. ~~Example of~~ Panel a~~weekly time series~~): Monthly average of ~~vertical profiles referred to lovbio035b BGC-Argo float~~ DCM depth for west (~~Figure 4~~blue) ~~showing REF simulation~~ and ~~alternative bio-optical models OPT1 and OPT2 compared to BGC-Argo float Chl values~~ east (~~thicker line~~orange) profiles derived from REF simulation. ~~The horizontal dashed blue line represents the euphotic depth z_{eu}~~ Circles and crosses are mean, ~~whereas~~ triangles are medians. Panel b): Scatter plots of the ~~dashed black line indicates the depth where~~ residual difference between measured ~~PAR equals 0.5 as identified in ?~~and modeled DCM. The ~~legend~~ x-axis reports model configurations listed in Table 1. On the y-axis, residuals' median values for west (blue) and east (orange) profiles are shown. Triangles indicate the 25th and 75th percentiles. Panel c): K_d for west and east subbasin during stratified period, diamonds indicate the median over the vertical column, 25th and 75th percentiles are the horizontal lines. Crosses show the maximum over the vertical column. Panel d) is the same as c) but with double initial nutrient concentrations for the western basin simulations.

~~Panel a): Scatter plots of the residual difference between measured and modeled DCMs. The x-axis reports model configurations listed in Table 1. On the y-axis, residuals' median values for west (blue) and east (orange) profiles are shown. Triangles indicate the 25th and 75th percentiles. Panel b): K_d for west and east subbasin during stratified period, diamonds indicate the median over the vertical column, 25th and 75th percentiles are the horizontal lines. Crosses show the maximum over the vertical column. Panel c) is the same as b) but with double initial nutrient concentrations for the western basin simulations.~~

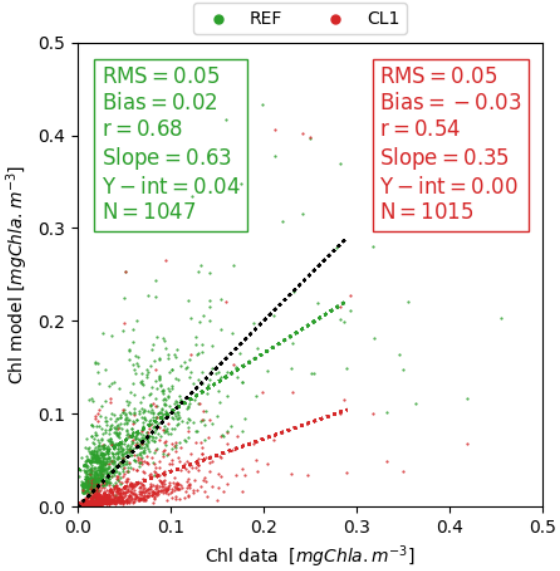

Figure 10. Scatter plot comparing 0-25 m average surface Chl versus BGC-Argo float data for the stratified period condition (DCM > 40 m).

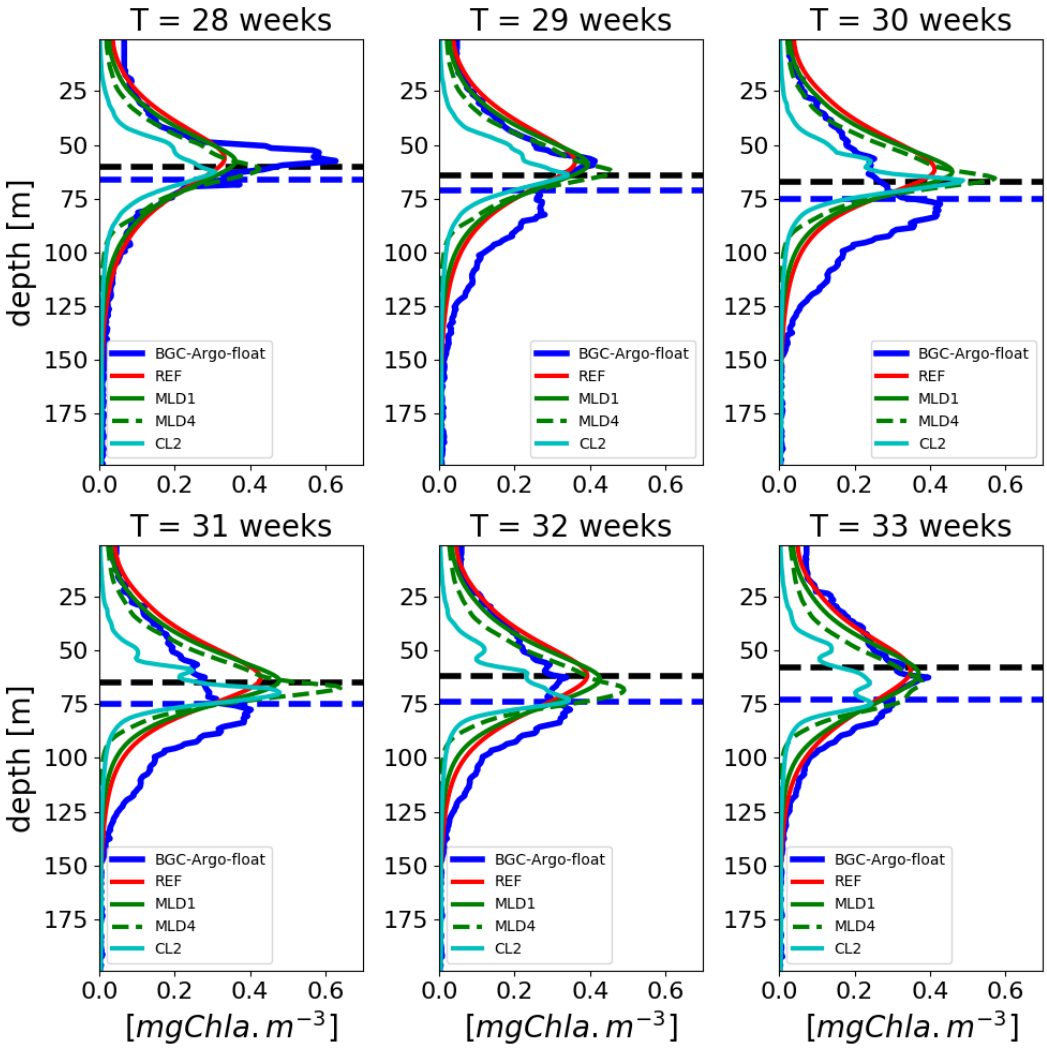

Figure 11. Example of a weekly time series of vertical profiles referred to lovbio035b BGC-Argo float (~~Figure~~ Fig. 4, from week 28 to week 33) based on diel variability and constant daily light descriptions, compared to BGC-Argo float Chl values (thicker blue line). The horizontal dashed blue line represents the euphotic depth z_{eu}, whereas the dashed black line indicates the depth where measured PAR equals 0.5 $\mathrm{molquanta\,m^{-2}\,day^{-1}}$ as identified in **?**. The legend reports model configurations listed in Table 1.

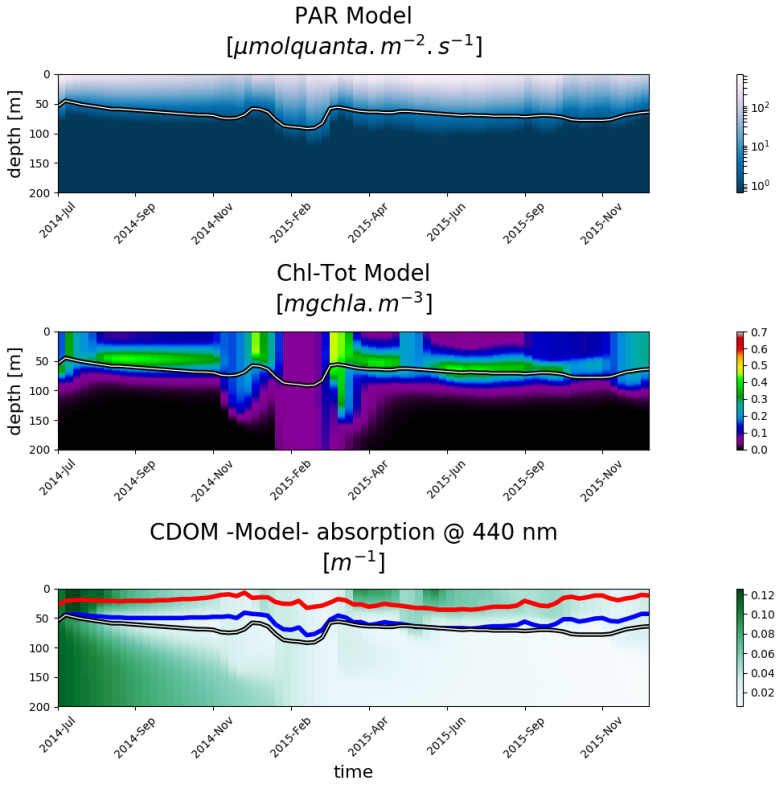

Figure 12. Hovmöller diagrams for BGC-Argo float lovbio068d (WMO code 6901648) deployed in the North West Mediterranean showing: PAR (top), total ~~chlorophyll~~ Chl (middle) and CDOM (bottom) simulated by model configuration OPT5. The white, red and blue lines depict the euphotic, 100% and 10% bleaching depths respectively.