Report #1

We thank the Reviewer for providing feedback. We propose to modify the manuscript according to the comments of the two reviewers as outlined in point-by-point replies. In bold our responses, in blue the actions.

General comments:

There are some spelling and grammatical errors throughout, and inconsistent use of tenses etc. We believe the journal offers copy editing as standard, so correcting these could potentially wait for that stage – we leave that decision to the editor.

Thank you for your comment. We have carefully revised the text, paying attention to grammar errors and inconsistent use of tenses, also integrating the Rev#2 typo corrections. If the editor suggests an additional copy editing check at this stage, we will promptly provide it

Use of sub-basin names is inconsistent (abbreviated or not; e.g. in Section 3.3: Nwm v North Western Mediterranean; eastern sub-basin v ion2); inconsistent use of terms like “reconstructed nitrate profiles” v “recNO3” (e.g. around L327) or names of model runs (e.g. L372 “reference run” which probably refers to).

We appreciate the reviewer's comments.

-Regarding the division of the Mediterranean basin into 16 sub-basins, we have decided to identify each with lowercase acronyms. Conversely, the 6 aggregated macro basins are represented with capitalized letters. One of the 16 sub-basins coincides with one of the 6 macro basins (i.e., the "North Western Mediterranean”). Thus, it is named nwm and Nwm when referred to as one of the 16 sub-basins or 6 macro basins, respectively. To avoid confusion in section 3.3, we have modified the text and title of figures using only the "nwm" acronym.

-"recNO3" and "reconstructed nitrate profiles" are interchangeable terms. We have refined the phrasing for clarity (new text is underlined):

“A generalized slight worsening in the assimilated runs can generally be observed during the summer stratification period and especially the Eastern sub-basins. From DAfl to DAnn, the value of RMSE slightly increases in all sub-basins. These values correspond to an average worsening of about 6% in DAfl and 7.5% in DAnn compared to the HIND run.

Despite the introduction of a significant number of reconstructed nitrate profiles in some sub-basins (e.g., orange striped lines of nwm and ion2 in Figure 3), this inclusion does not positively impact the summer chlorophyll RMSE at the surface.”

- We have replaced “reference run” with HIND: “Differences between the assimilation and the HIND run accumulate over time”

Introduction: For readability and clarity, it would be useful to link the different topics better and state what the gaps and advantages are that you are addressing in the results with the modular approach. The motivation is not clear from the introduction. For example: the transition between DA and NN in the introduction (L65) could be done by stating which gap NN can fill for the DA, i.e. adding reconstructed observations which improving the DA

analysis depends on.
Thank you for your comment. Here's the rearranged new text:
"**In recent years, data assimilation (DA) techniques have increasingly incorporated neural network (NN)-based tools. The main strength of NN algorithms lies in their ability to approximate continuous functions (Hornik et al., 1989) in remarkably low computational times. These NN-based tools have been integrated into DA frameworks to tackle various DA challenges, such as bias correction (Kumar et al., 2015; Zhou et al., 2021), reformulation of observation operators (Storto et al., 2021), and cross-calibration (Lary et al., 2018). Furthermore, NN algorithms are frequently used as independent tools, distinct from data assimilation, for generating new products and/or reconstructing datasets (Lary et al., 2018). The use of reconstructed datasets may compensate for potential gaps in observation availability, potentially enhancing the predictive skill of numerical models".**

Introduction: Paragraph about the evolution of MedBFM (L83ff) is very nice now.
Thank you for the very positive feedback.

Results: From Section 3.3, you either introduce or summarise the approach taken to show the results of the analysis, which greatly helps the readability of these Sections. It would be great if you could add similar introductions to 3.1 and 3.2, to give an overview of which variables are assessed and how, etc.
Thank you for your feedback.

Section 3.1: no changes will be made to the text.
After careful consideration, we believe that Section 3.1 effectively presents the necessary information and thus, we would prefer to keep the text as it is. Specifically, in lines 285-291, we have provided comprehensive statistics regarding the QC O2 module, also comparing our average value with values from literature. Furthermore, in lines 292-295, we have explained the implications of this correction on a single float, and in lines 296-298, we have discussed a collateral effect derived from our approach. We are confident that these sections sufficiently address the relevant aspects of our study.

Section 3.2 has been rephrased as follows, L314-315 has been canceled:
"Skill performances of the simulations listed in Table 1 are evaluated by comparing model results with (i) the satellite Marine Copernicus OC product (i.e., non-gap-filled L3 product OCEANCOLOUR_MED_BGC_L3_MY_009_143 from marine.copernicus.eu, last visited in July 2023) of chlorophyll and (ii) BGC-Argo profiles of chlorophyll, nitrate, and oxygen (Argo, 2022). The satellite OC L3 products downloaded from the Copernicus Marine Service catalogue are interpolated from 1 km to the 1/24° model resolution.

Specifically, we compared the daily model output with the satellite dataset and the model's first guess (i.e., the model state at 1pm before assimilation) with the BGC-Argo profiles. While the use of the first guess is a common practice in data assimilation (Hollingsworth et al., 1986), it is worth to remind that this comparison should be considered as a semi-independent validation, given that two consecutive profiles of the same BGC-Argo float can share a certain degree of correlation in their errors.

The Root Mean Square Error (RMSE) metric is chosen to quantify the model capability to reproduce seasonal variability of the main biogeochemical (BGC) processes at the surface (satellite dataset) or along the vertical column (BGC Argo dataset), such as phytoplankton surface bloom and dynamics during water column stratification.

Indeed, the RMSE is evaluated during winter (from February to April, FMA) and summer (from June to August, JJA) 2017 and 2018 within 16 sub-basins of the Mediterranean Sea (as described in Section 2.4 and in Figure 2) or in an aggregated combination of them."

Technical comments:

L49-53 The info in the bracket (L50) interrupts the reading flow and may merit its own sentence. The information is also partially repeated in L52 (">1% per year").
Thank you for your feedback. We have revised the text as follows:
"Despite efforts to correct drift during storage, which may enhance accuracy by 5-10%, it is likely that an in situ (or during deployment) drift is still observed. For instance, Maurer et al. (2021) observed significant drift rates in about 25% of the 126 floats analyzed for the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) project. These drift rates spanned a total range of -1.1 to 1.2% per year, with a standard deviation of 0.65% per year. Similarly, Bushinsky et al. (2016) found the presence of significant drift rates in about 70% of the floats deployed in the Northern Pacific Ocean. Notably, both positive and negative drift rates were observed across various studies, including those by Johnson and Claustre (2016), Bushinsky et al. (2016) , Bittig et al. (2018b) and Maurer et al. (2021)."

L184 "inconsistencies between the deeper (below 600 m) and the lower part of the assimilated layer." It took a few reads to grasp the distinction between "deeper" and "lower" here, suggest rewording for clarity.
Thank you for your feedback. We have revised the text as follows:
"This adjustment aims to prevent inconsistencies between the lower part of the assimilated layer (450-600m) and the deeper layer of the water column (below 600m)."

L202 "EMODnet" – It would be worth adding one or two sentences describing that data set for those unfamiliar with it, e.g. what data it is based on, if it is gridded or not, etc.
We have revised the text as follows:
"The NN-MLP-MED introduces several innovative features compared to the mentioned methods (e.g., CANYON-Med; Fourrier et al. 2020) leading to improved results.

Firstly, the input dataset encompasses a larger sample size and broader coverage of the Mediterranean Sea region. The EMODnet (European Marine Observation and Data Network) data collection, as described by Buga et al. (2018), consists of multi-platform data gathered from different research cruises and monitoring activities in Europe's marine waters and global oceans. This dataset is characterized by its multivariate nature, including various biogeochemical observations such as chlorophyll, nitrate, phosphate, dissolved oxygen, DIC, and alkalinity, collected between 1999 and 2018. Additionally, this dataset is further enriched with in situ observations spanning the

period from 1999 to 2016, as detailed in Lazzari et al. (2016) and Cossarini et al. (2015b).

L214 “a balanced distribution” – in Fig. 3 it looks like there are more summer profiles added than winter profiles rather than similar numbers of profiles in both seasons.
We have revised the text as follows (new text underlined):
“After incorporating the reconstructed profiles (recNO3), the nitrate dataset used for assimilation expands to 2146 profiles from the initial 938 nitrate (NO3) profiles (Table 1). Generated by the NN-MLP-MED module, the reconstructed dataset offers broad spatial coverage across the 16 regions of the Mediterranean Sea (Figure 2), as well as a quite balanced distribution of nitrate data throughout the seasons (Figure 3), with the addition of 218 reconstructed profiles of nitrate in winter and 361 in summer, respectively.”

L248 and elsewhere: “mmol m−3 y” – should this not be “mmol m−3 y-1”?
Thank you , we have corrected it with: mmol m−3 y−1.

L249-252 “linearly interpolating” – what is the basis for that? Is there a reference saying the drift has a linear dependence on depth/pressure? “where drift is set equal to zero.” – Is this the result of QC at the DACs (mentioned in the following sentence)? These sentences are worth clarifying, in particular, which aspect is from the literature and which is an assumption made in this paper. If drift actually changes non-linearly with depth, then the correction could be introducing a source of error – this should be discussed.
Thank you. As detailed in line 252 “The presence of near-surface tests motivates our decision to mitigate the correction’s impact at the surface” rather than a “linear dependence on depth/pressure” (which we have not mentioned in the text). GDACs perform more than 14 tests before releasing data in AM and DM based on oxygen concentration at the surface. Conversely, specific tests for correction based on oxygen concentrations at depth have not yet been developed by the DACs.
Our methodology has been thoroughly discussed in several meetings (e.g., the 24th Argo Data Management Team Meeting in Hobart from October 23-27, 2023) and has been favorably welcomed by the BGC-Argo community. Based on the aforementioned motivations we would not introduce any changes to the text.

L261 “and the initial conditions of oxygen which are retrieved from BGC-Argo float climatology computed after QC O2 procedure” – the initial conditions must be on the model grid. How is this achieved including QC O2 and where is the BGC-Argo climatology coming from?
Thank you. We have added the required information as follows:
"[...] and the initial oxygen conditions. These conditions are derived from the BGC-Argo dataset by generating 16 climatological profiles of oxygen after performing the QC O2 procedure, and then uniformly assigning them to each grid point of the 16 sub-basins shown in Figure 2."

L294 “After 2 years, the bias due to the drift reaches...” – As I understand it you perform a drift correction on a profile-to-profile basis. It may be worth stating somewhere if the drift is linear over time?
As we have learnt at the meetings with researchers from different GDACs, the in situ drift typically reaches its maximum after about one year from the first deployment

(with a drift approximately 1%) and the rate of drift remains almost stable from the second year onward. Therefore, we have implemented the criteria that the drift is calculated only if the timeseries is longer than 1 year (L242-243)

L294 How does the drift behave from one profile to the next and long-term over time?
The following figure (R1 left) shows the time series of drift values at 600m calculated for the BGC-Argo float 6901765, mainly located in the Aegean Sea (aeg) and the Ionian Sea (ion2), with a few measurements in the Levantine (lev1). The figure R1 (left) exhibits some small oscillations during the initial period and a convergence to 2.5 mmol/m3/y after 2 years from the deployment (around 2017-05). The standard deviation of the 2017-2018 timeseries of the drift rate is 6% (0.15 mmol/m3/y) which is reasonably low compared to the average value of the drift.

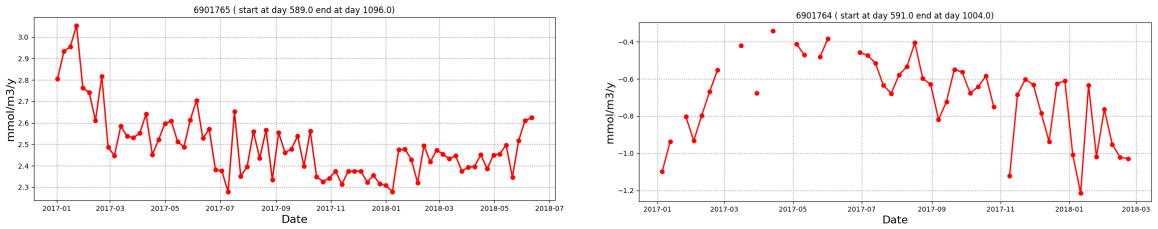

Figure R1. Time series of the drift rate (mmolO2/m3/y) for the BGC-Argo floats 6901765 (left) and 6901764 (right). Drift rate is shown for a given BGC-Argo timeseries starting from 1 Jan 2017 which is more than 1 year after the float deployment (i.e. March 2015 for both floats).

L294 Will the drift continuously exceed the threshold after one profile exceeded it?
Generally, yes. A nice example is shown in figure R1 left (see previous comment). However, we found a few exceptions (less than 1% of the profiles in the 2017-2018 dataset). These occur when the drift rate is very close to the threshold (1mmol/m3/y). One of the few examples can be seen in Figure R1 (right), where the drift at 600 m fluctuates below and above the threshold four times in the period from November 2017 to March 2018. In these few cases, when applied, the correction is small.
no changes will be made to the text.

L303-306 Thank you for adding this clarification, but it might be worth specifying “a certain degree of correlation in their errors.”
Thank you for suggesting a more accurate wording, we have corrected the sentence as you proposed.

L315 “a composite weekly average was computed to ensure gap-free maps” – in your response to our previous review you stated “The weekly was a typo, we actually used the daily L3 map of satellite chlorophyll from Copernicus. They are given as daily maps thus the comparison uses the model as daily output.” Please modify the text if this is the case, and also clarify if you used a L3 (non-gap-filled) or L4 (gap-filled) product.
Thank you, we have modified the paragraphs and corrected the inconsistency. The

new version of the paragraph is proposed on page n.1 of this document (Skill performances..).

L325: “RMSE” – do you mean RMSE reduction?
L325: “which increases in all sub-basins” – is “which” referring to the RMSE or the chlorophyll?
Thank you for the comments. We have revised L325 as follows:
"From DAfl to DAnn, the value of RMSE slightly increases in all sub-basins. These values correspond to an average worsening of about 6% in DAfl and 7.5% in DAnn compared to the HIND run."

L327: “reconstructed nitrates” – reconstructed nitrate profiles
L328 What do you mean by “shallow statistics”?
Thank you for the feedback. We have revised L327-L328 as follows:
"Despite the introduction of a significant number of reconstructed nitrate profiles in some sub-basins (e.g., depicted by the orange striped lines of nwm and ion2 in Figure 3), this inclusion does not positively impact the summer chlorophyll RMSE at the surface."

L345-348 Fig 6 middle panel does not show any large improvements in the chl statistics, regardless of assimilating chlorophyll profiles or adding more nutrient profiles.
As explained in L333 “The statistics computed over the aggregate basin provide more robust results (e.g., they are computed over a larger number of profiles) even if possible spatial patterns of the errors can be damped. Thus, this choice might limit the analysis on whether/how different nitrate assimilation setups affect chlorophyll and oxygen dynamics (see Section 3.3).”
The purpose of plotting seasonal RMSE in aggregated sub-basins is to demonstrate that the assimilation of reconstructed nitrate profiles does not diminish the model's skill to reproduce bloom and stratification BGC-dynamics. Figures 7-13 aim to explain the enlarged impact experienced by the different DA setups. Based on the aforementioned motivations we would not introduce any changes to the text.

L351-353 “As discussed in Section 2.2” – In Section 2.2 you described how the oxygen variability and oxygen assimilation does not strongly affect the wider BGC, but here you imply that this means assimilating nitrate does not affect the oxygen strongly, which is the opposite argument. Also relevant to lines 449-451.
Thanks for the comment. Given that O2 profiles are assimilated at the same location of the NN-nitrate profile assimilated, it is not expected and observed any difference between DAnn and DAfl in terms of oxygen.
We have rectified the sentence at line 351-353 as follows:
The integration of reconstructed profiles in the DAnn simulation does not significantly affect oxygen dynamics compared to the DAfl simulation, given that oxygen has already been markedly modified by the O2 assimilation occurring at the same location as nitrate NN-reconstructed profiles. Additionally, lines 449-451 have been corrected as follows (new text underlined): Oxygen impact maps (not shown) are very similar to the nitrate DAnn maps and do not show significant differences between the two DA simulations, <u>since the same QC oxygen dataset was assimilated in DAfl and DAnn and the oxygen assimilation largely overcome any other potential model adjustment after nitrate assimilation.</u>

L368 “corrects” implies that the bias disappears completely, while the following sentence says there is more correction in Dann. It may be better to say the assimilation “reduces the bias” or something similar.
ok, we have replaced the unclear use of “corrects” in the sentence with “reduces a general positive bias”

L405-409 Unclear use of statistical terms: 10% change is “most significant” but 5% change is “negligible”. Can you rephrase e.g. using comparative words (like smaller, larger) rather than statistical terms (significant, negligible) please. Also relevant elsewhere in the manuscript.
Thank you for bringing this issue to our attention. We have corrected as follows:
“In the DAfl simulation, the most evident differences in primary production compared to the HIND simulation are located in the Eastern Mediterranean Sea with a decrease of NPP of nearly 10% in the Levantine macro-basin and in the Ionian Sea close to the Greek coast (first and second row of Figure 11). This reduction is particularly pronounced during winter. In the Western Mediterranean the impacts on primary production are less evident in both seasons with a slight reduction (5%) in winter in the Tyrrhenian Sea.”
Use of "significant" or "negligible" has been carefully revised throughout the text.

L414-416 The phrasing of this paragraph is ambiguous and took a few reads to be clear of the meaning. Please rephrase for clarity.
We have modified the text as follows:
“As shown in Figure 3, basins lev1 and lev4 have a high number of reconstructed nitrate profiles during both winter and summer seasons. This abundance of reconstructed profiles contributes to an increase in impact in reproducing the NNP dynamics, which is spatially localized. Conversely, lev2 and lev3 the sub-basins dividing basins lev1 from lev4, contain in situ nitrate and lack of reconstructed nitrate profiles. This lack may spatially limit the impacts that assimilating reconstructed nitrate profiles could have on NPP throughout the entire Levantine region (Lev).”

L435 Is the value of the 95th percentile (i.e. 0.1) different between DAfl and DAnn? If so, doesn’t the impact parameter mix the area impacted by the DA as well as the magnitude of the changes?

Thank you for raising the question about the threshold on I(t). This will help to clarify this aspect. Indeed, the “95th percentile” refers to the impact indicator I(t) for each period, variable and simulation, thus it is not a fixed value but it provides a map (for each period, simulation and variable; Fig. 11 and 12). Being a map that describes the areas with the “largest” (95th percentile) relative differences between DAfl or DAnn and HIND, it is consistent to compare I(t) 95th percentile for each variable and different simulations. It is also worth noting that the 0.1 threshold has been used only to give the reader a visual reference for comparing the maps, whilst the 0.1 value does not play any role in the map calculation. The value of 0.1 is calculated after merging all the DAnn and DAfl values of the I(t) 95th percentile into a unified set of data.
We will clarify this aspect as follows:

In DAfl, the extent of nitrate Iij (t) 95th above 0.1 (which represents the mean of the 95th percentile impact indicator in the Mediterranean Sea calculated after merging all the DAnn and DAfl Iij (t) 95th values) is 16.5% and 18.7% in winter and in summer respectively, with a clear spatial distribution mapping the density of BGC-Argo floats.

L445 “impact to almost all the Mediterranean Sea” – A few sentences before you state that the impacted area increases to about 30% when including the reconstructed profiles. How do you conclude from this that the approach has the ability to encompass “almost all” of the MedSea?
Thank you, we referred to the number of sub-basins (over 16) involved in this spatial impact and not to the % of impact. we have rectified as follows:
“These results suggest that the inclusion of reconstructed nitrate assimilation has the potential to extend its impact across the majority of the 16 sub-basins of the Mediterranean Sea. However, the scarcity or absence of available data for assimilation prevents us from observing an impact in the marginal seas (Adr and Aeg), the southern part of the Ionian (ion1), and Western sub-basins (alb and swm1).”

L483 “in each 2.5deg x 2.5deg box every 10 days for the 2017-2018 period” – Is this estimate of one float per 2.5x2.5 deg an average of all available floats over the MedSea area? Or a theoretical aim? The distribution of the measured and reconstructed profiles is highly heterogeneous. Wouldn’t that affect the necessary number of floats to constrain the BGC? And is that number of one float per 2.5x2.5 deg to constrain the BGC your hypothesis or a result of a previous study (e.g. an OSSE)?
It is an estimate coming from the resolution of our dataset rather than a theoretical aim or a result of a previous study. Our results show a high level of the impact when the density of float is higher to the proposed number. In fact we conclude that the uneven distribution of the BGC-Argo float allows that the "mesoscale dynamics can only be locally constrained”. For instance, within our BGC dataset, it may be feasible to study the mesoscale dynamics of the nwm sub-basin, while, as noted in Line 494, some sub-basins (alb, Adr, Ion1, and Aegean) are still under-sampled. no changes will be made to the text.

L486-487 “a further increase of the area impacted from a float assimilation can be achieved by redefining horizontal covariance errors” – such an increase is only desirable if the correlations are real, otherwise the increased “impact” may actual degrade rather than improve the analysis. It would be better to talk about “optimising” this, which would better link with the next sentence.
Thanks, we agree and reformulate as follows:
"Apart from an increase in the numbers of floats, a further increase of the area impacted from a float assimilation can be optimized …”

L501 “a validation error of 0.50 mmol m−3 for nitrate and 0.87 mmol m−3 when applied to predict BGC-Argo data.” For clarity, please rephrase to something like “a validation error of 0.50 mmol m−3 when used to predict nitrate from the EMODnet data set, and 0.87 mmol m−3 when used to predict nitrate from BGC-Argo data.”
Thanks, we agree and reformulate as follows:
“The MLP-NN-MED method exhibits a validation error of 0.50 mmol m−3 for nitrate when used to predict nitrate from the EMODnet data set, and 0.87 mmol m−3 when used to predict nitrate from BGC-Argo data (Pietropolli et al., 2023).”

L504 “Using the same error for both datasets revealed the highest potential impact of the reconstructed nitrate.” If the measure of “impact” is best matching the assimilated observations (which is what this sentence implies), then the “highest” impact would be shown by using zero observation error. Please rephrase this sentence.
Thanks, we agree and reformulate as follows:
“**Thus, while it is reasonable to assign a higher observation error to NN reconstructed nitrate, applying the same error to both in situ and NN reconstructed datasets has resulted in a potential overestimation of the assimilation impact that can be achieved.”**

References: Bittig et al. 2018a and b are identical; spelling of “d’Ortenzio” or “D’ortenzio” in the two references; Vichi et al. 2007 a and b are identical.
Thank you, we have corrected the reference inconsistencies.

Report #2
We thank the Reviewer for providing feedback. We propose to modify the manuscript according to the comments of the two reviewers as outlined in point-by-point replies. In bold our responses, in blue the actions.

General comments:

Authors answered the reviewer's requests and questions thoroughly and major issues in the previous version of this manuscript were resolved. The improved quality of figures make this article easy to understand. Additional sentences and paragraphs, especially in introduction, and further detail on NN-MLP-MED in section 2 made the objectives of this study and article clearer. However, these additional sentences introduced additional ambiguities and editorial issues at the same time. This article could be published after making some minor corrections as suggested below.
Thank you for your feedback, which greatly contributed to enhancing the readability of the manuscript. We are pleased to hear that the major issues identified in the previous version have been effectively addressed. We will carefully address the minor corrections you have suggested.

Scientific/Technical questions and issues:

P4.L104: "Because of its particular characteristics" Not sure what does it mean by term "particular characteristics". If it means what are described in the paragraph.
P4.L107-P4.L119, which part of the characteristics makes the Mediterranean Sea ideal site of the OSE? For example, the presence of season- and domain-dependent DCM and nitracline depths attracts the idea of assimilating BGC profiles since they are not observable from space and commonly subject to relatively large model bias or representativeness error. Please be more specific about this point.
Thank you for your feedback. Yes, we used the sentence "its particular characteristics," to introduce the following paragraph. However, we have rephrased and added information to emphasize that all the characteristics listed for the Mediterranean Sea were equally important to support the choice of the Mediterranean Sea as area of study.

“Given its characterization as a miniature ocean suitable for climate studies and considering the density of BGC-Argo profiles, the Mediterranean Sea represents an ideal site for conducting Observing System Experiment (OSE) studies to assess the feasibility of assimilating BGC-profiles and analyzing their impacts.

Indeed the Mediterranean Sea is an anti-estuarine semi-enclosed sea (Pinardi et al., 2015) with a complex overturning circulation. This circulation consists of horizontal mesoscale and sub-basins scale gyre structures, transitional cyclonic and anticyclonic gyres and eddies. These dynamics are influenced by bathymetric features interconnected by currents and jets (Oddo et al., 2009), along with vigorous vertical velocities. Furthermore, the shallow Sicily Strait, with a depth of approximately 500 meters, separates the Western Mediterranean from the Eastern Mediterranean. This geographical feature allows different processes to dominate in each of the two regions and limits exchanges only between surface and intermediate waters (Pinardi et al., 2015).

Even from a biogeochemical (BGC) perspective, the Mediterranean Sea can be roughly subdivided into the Western and Eastern Mediterranean sectors, characterized by an oligotrophic West-East gradient. This gradient results in low nutrient availability at the surface, which is generally insufficient to sustain high phytoplankton biomass (Siokou et al., 2010; Marañón et al., 2021). Additionally, there is a deeper nitracline in the east (>120m) compared to the west (<100m).

Chlorophyll [.....] (Dibiagio et al., 2022).

While the general dynamics of biogeochemical processes can be summarized in a two-basin gradient, it's important to note that mesoscale and sub-mesoscale events can significantly impact the Mediterranean Sea at the sub-basin scale. These events can create intense local dynamics, such as, such as blooms and water column stratification, which are often associated with eddy activities and peculiar vertical circulation. Reproducing these phenomena in numerical model simulations can be more challenging, as they are prone to encountering high model bias or representativeness error.”

P5.L143-L145: "OGSTM .. it is forced by the output (..) of the NEMO3.2 model .." According to this paragraph, OGSTM solves tracer equations off-line with the output of the NEMO3.2 model. However, BGC tracer equations in BFM require external atmospheric forcing such as PAR. Can you describe the external forcing here?
Thank you, we added the information as follows:
OGSTM solves for advection, diffusion, sinking terms, and considers the effects of the free surface and variable volume-layer effects on tracer transport (Salon et al., 2019). It is forced by output variables such as current, temperature (T), salinity (S), and sea surface height from the NEMO3.6 model (Clementi et al., 2017). OGSTM and NEMO3.6 share the same bathymetry and z*grid configuration, as well as open boundary and river conditions (Coppini et al., 2023). Atmospheric forcing, including solar shortwave irradiance and wind stress, is acquired as 2-D daily fields from the European Centre for Medium-Range WeatherForecasts (ECMWF), as detailed by Salon et al. (2019).

P6.L156: ".. which relies on the misfit between the model background (xb) and the observations (y) .." This statement and the equation (1) are not correct. I am sorry to miss this mistake in the first review comments. Cost function of 3DVar is weighted sum of two terms, 1) a misfit between model control state variable (xa) and its background estimation (xb) and 2) a misfit between observations (y) and its model correspondent (H(xa)). Please fix the description and equation (1).
Thank you for highlighting the mistake in the definition. Here the new version:
"This function comprises two terms: (i) the misfit between the model background (xb) and the model control state variable, or analysis (i.e., the assimilation result xa) and (ii) the mismatch between the observations (y) and the analysis (xa). Both terms are weighted by their respective error covariance matrices (B and R) as follows:
$$J(xa) = (xa - xb)^T B^{-1}(xa - xb) + (y - H(xa))^T R^{-1}(y - H(xa))$$"

P7.L199: "from temperature and salinity (Argo), oxygen (BGC Argo) and float date .." As far as I understand, three input variables (temperature, salinity and oxygen) and coordinate information (date, lat and lon) are all from the same BGC Argo profile. If it were the case, this statement is misleading. Should you state simply " from sets of temperature, salinity, oxygen, date, latitude and longitude of the BGC Argo profiles."?.
Yes data came from the same float. We have corrected as follows:
"In our OSE experiment, the trained NN-MLP-MED reconstructs nitrate profiles from sets of temperature, salinity, oxygen, date, latitude and longitude BGC-Argo profiles."

P14.L346-L348: "This is because the direct ..." This statement is still speculative. As far as I read this article, there is no evidence supporting this statement. The OSE experiment is not designed to measure size of impact of chlorophyll-a assimilation and nitrate assimilation independently to chlorophyll-a profile analysis. Plus, comparison with HIND indicates that assimilation of chlorophyll-a profile itself is not effective to reduce profile chlorophyll RMSE in most of the area except for Lev. I suggest removing this statement.
we agree and have decided to remove the statement

Editorial issues:

P2. L49-L51: .. a drift in about 25% (..) and 70% of analyzed floats, respectively. Not clear what differences were found in 25% and 70% of analyzed floats, respectively.

Following all the reviewers' comments, we have added information and rephrased the paragraph as follows:
"Despite efforts to correct drift during storage, which may enhance accuracy by 5-10%, it is likely that an in situ (or during deployment) drift is still observed. For instance, Maurer et al. (2021) observed significant drift rates in about 25% of the 126 floats analyzed for the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) project. These drift rates spanned a total range of -1.1 to 1.2% per year, with a standard deviation of 0.65% per year. Similarly, Bushinsky et al. (2016) found the presence of significant drift rates in about 70% of the floats deployed in the Northern Pacific Ocean. Notably, both positive and negative drift rates were observed across various studies, including those by Johnson and Claustre (2016), Bushinsky et al. (2016) , Bittig et al. (2018b) and Maurer et al. (2021)."

P3. L63: ", and solving problems .." > ", and solve problems .." →**OK**

P3.L78, P8.L219: "Canyon-b" > "CANYON-B" →**OK**

P7.L201: "Canyon-Med" > "CANYON-MED" →**OK**

P4.L99: " oxygen BGC-Argo profiles " > " BGC-Argo oxygen profiles" →**OK**

P4.L103: " BGC-Argo chlorophyll, nitrate, and oxygen" > " BGC-Argo chlorophyll, nitrate, and oxygen profiles" →**OK**

P4.L104: "in situ observations" > "the in situ observations" or "the BGC-Argo profiles" or "the BGC profiles" →**OK**

P4.L104: "reconstructed ones" > "NN reconstructed profiles" for clarity →**OK**

P5.L127: "reconstructed profiles " > "NN reconstructed profiles" for clarity →**OK**

P5.Figure 1. The term "OGSTSM-BFM" appears here for the first time and "OGSTM" and "BFM"are described for the first time in subsection 2.1.
We enlarged the paragraph introducing the OGSTSM-BFM acronym.
"In the following sections, we introduce the components of the MedBFM system, including the transport model (OGSTM, Foujols et al., 2000; Lazzari et al., 2012; and Lazzari et al., 2016) and the biogeochemical flux model (BFM, Vichi et al., 2007a; Vichi et al., 2007b). Additionally, we describe the novel modules, namely the QC O2 procedure and the NN-MLP-MED scheme. Furthermore, we outline the dataset, which comprises BGC-Argo and NN reconstructed datasets, and discuss the revised 3DVarBio approach."

P5.L139: ".. versions, the BFM, Biogeochemical Flux Model .." > ".. versions of the Biogeochemical Flux Model (BFM) .." →**OK**

P6.L143: "the NEMO3.2 model" > This term appears for the first time here and needs proper citation or explanation. → **OK**

P7.L185: ".. we decided to not use .." > ".. we decided not to use .." →**OK**

P7.L185-L186: "in order to show the highest potential impact of the OSE." > Do you like to say "in order to show the highest potential impact of the NN reconstructed nitrate profiles to the OSE."? → **yes**

P7.L189: "(2002)" > "(2002)." →**OK**

P7.L200: "the mentioned methods" > Not clear which methods it is referring to. Does it refer to paragraph in P3.L75-L82?

"with respect to the previous CANYON's methods"

P8.L226: "sub -basins:" > "sub -basins (figure 2):"→**OK**

P8.L233: "All the three BGC variables" > Which three BGC variables it is referring to? A set of (recNO3, NO3 and Chl)? If it were the case, please state "All the three BGC variables (recNO3, NO3 and Chl)" to be more specific.
OK, we specified the variable in the text

P9 Figure 2 legend and caption: It is helpful for reader to be indicated that oxygen profiles are assimilated at the location of blue markers here as stated in P14.L353.

The distribution of oxygen profiles (not directly shown) can be inferred by examining the distribution of the blue dots.

P13 Figure 5 caption: "chlorophyll RMSE" > "OC chlorophyll RMSE" for clarity.→**OK**

P13.L328: What does it mean by "summer chlorophyll shallow statistics"?
Thank you for the feedback. We have revised L327-L328 as follows:
"Despite the introduction of a significant number of reconstructed nitrate profiles in some sub-basins (e.g., depicted by the orange striped lines of nwm and ion2 in Figure 3), this inclusion does not positively impact the summer chlorophyll RMSE at the surface."

P24.L502: "higher then the one" > "higher than the one" →**OK**