#REV 2

egusphere-2023-1588

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

General comments:

Scientific significance:

Reconstruction of nutrient information in CANYON-B based ANN system NN-MLP-MED relies on high accuracy of in situ O2 data while Argo on board O2 sensor is known to suffer significant sensor drift. To mitigate the Argo O2 data drift problem, authors have introduced QC O2 module for further calibration of O2 profile data. This novel approach in conducting secondary O2 calibration is a key component of this study.

Since the pioneering work of Ford et al. (2021), impact of sparsity of BGC Argo profile in ocean state estimation or data assimilation is recognised as a clear issue in BGC Argo profile data assimilation study and operational system. OSE experiment with and without NN-enlarged nitrate profile data for assimilation demonstrated that usefulness of machine-learning retrieved nutrient data can improve model representation of surface phytoplankton dynamics at certain conditions. Impact indicator study reveals clear impact of reconstructed nitrate profile assimilation in model BGC state, especially in the upper macronutrients and chlorophyll-a fields.

Scientific quality:

Scientific question raised in this study is clear and important one. Current density of BGC Argo floats array does not cover even basin scale ocean circulation which is an original target of CORE Argo float deployment goal. With the advancement of NN-base BGC variable retrieval methods, it is natural to test if such generated data can help us constrain ocean model for state estimation in operational settings. This study indicates positive impact of such data. However, many evaluation procedures to judge detail impact of the NN-derived nitrate data are not designed effectively to achieve its goal.

We appreciate the constructive comments and suggestions from the Reviewer. We present our point-by-point responses to the Reviewer's comments below. The Reviewer's comments are in blue, our responses follow each comment in black. In each response, we detail the changes we propose to make to the manuscript and include the proposed modified text and/or figure (in red).
For clarity, we have numbered some of the reviewers' comments so that similar ones are aggregated to provide a single response. Comments are labeled and highlighted with specific colors to distinguish reviewers (e.g. Rev. 1: **comment1a**, Rev. 2: **comment1b,** Rev. 3: → **comment1c**).

While the scientific contribution of this study is significant, there is clear problem in how it is delivered as journal paper. Overall, many statements are "speculative" or "subjective" for a data assimilation OSE study and many of statements are not supported directly by provided materials. Typical examples are presentation of RMSEs and difference between HIND and DA experiments in Figure 5 and Figures 6-9. Authors asked reader to read these number from figures rather than presenting actual numbers and the figures are generally presented not adequately for the purposes.
Thank you for the comment. We will take care of the readiness of Figures 5-9 by changing the font size and colors/textures.

comment1b
Authors should make clear differences between what can be concluded and what can be speculated from background knowledge. For example, authors discuss "impact of chlorophyll profile assimilation", but OSE setting has only Hindcast (HIND), DA w CHL, O2, NO3 profiles (DAfl) and DA w/ CHL, O2, NO3 profiles plus nn-derived NO3 profiles (DAnn). How can you discuss sole impact of chlorophyll assimilation with this OSE setting? Other cases can be found in the comments under "Specific comments".
We do agree, we will review the text avoiding speculative comments (see also **comment8b**). Here an example:
At Line L274, the positive "impact of chlorophyll profile assimilation" was presented in Teruzzi et al. (2021). In this work, "the impact of chlorophyll profile assimilation" can be assessed by comparing the HIND run with the other two DA setups (Figure 4 and 5). Figure 4 aims to show "*the positive impact that reconstructed nitrate profiles have on phytoplankton at the surface*". Figure 5 aims to show (i) "the positive impact obtained comparing the HIND to both the DA setups" and (ii) "*that chlorophyll assimilation is more effective than dynamical model adjustment after reconstructed nitrate assimilation".*
We will rewrite this sentence as:
L274: Since the DAfl and DAnn simulations share the same chlorophyll assimilative setup, the RMSE improvements with respect to the chlorophyll assimilation can be evaluated by comparing the HIND with DAfl or DAnn simulations (Figure 5 middle panel). For chlorophyll, a reduction of RMSE in DAfl (Figure 5 middle panel) is observed in nwm, ion and lev in winter and at depth in tyr, ion and lev in summer, which is in line to what already shown for nitrate and chlorophyll profiles assimilation in Teruzzi et al. (2021). The positive impact of DAnn on nitrate RMSE (Figure 5 top panel) is not directly transferred to the vertical chlorophyll RMSE (as observed for figure 4). This is due to the fact that direct chlorophyll assimilation is more effective than the dynamical model adjustment after nitrate and reconstructed nitrate assimilation in the areas close to the observed chlorophyll profiles.

Presentation quality:

While scientific value of this study is high, its presentation quality is rather poor as demonstrated in the long list of comments under “Technical corrections”. In general, size of figures and fonts are too small in most of figures. Choice of color scheme in Figure 4 and 5 is questionable for people with color vision deficiencies. Please see guideline on the preparation of graphs: https://www.biogeosciences.net/submission.html. There are many editorial issues ranging from simple wording issue from more serious structural issues. Further detail can be found under “Technical comments”. Combining issues raised in “Scientific quality” and “Presentation quality”, I recommend major revision. Comments follow below.

Thank you for the comment. We will redo all figures by enlarging the size of the text and when required changing the palette/texture. Proposed changes are detailed in the line-by-line replies to the "Technical Comments". The colors in Figure 4-5 were chosen after testing their effectiveness at the website: Coblis — Color Blindness Simulator. Since the chosen palette could be problematic for the “Green-Blind/Deuteranopia test”, we applied different marker styles to the lines in Figure 5. In Figure 4 the legend is in order of appearance of the bars.

Specific comments:

P2.l48: model tuning (Wang et al., 2020)

Please add more recent references on this topic:

Yumruktepe et al. 2023 https://gmd.copernicus.org/preprints/gmd-2023-25/

Wang and Fennel 2023 https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022GL101220

Thank you, we will add the references.

P6.l146: “VH is built using a Gaussian filter whose correlation radius modulates the smoothing intensity”

> What is the size of correlation radius in average? This information is important to understand how far BGC Argo profile assimilation leave impact in the analysis.

Thank you for the suggestion. The correlation radius ranges between 12-20 km. Detailed information on the tuning of correlation radius can be found in Figure 3 of Cossarini et al., 2019. We propose to clarify the point with the following sentence:

As in Cossarini et al. (2019), in this work the correlation radius is non-uniform, direction-dependent, and ranges between 12 and 20 km (16 km on average).

comment2b

P7.l184: “Adjusted and delayed mode data were selected for oxygen and chlorophyll”.

> Can you describe which QC flag was used for selecting “good” data both for oxygen and chlorophyll?

The information will be added also considering other comment (**comment14a**) at old L184 as follows (in bold the new text):

We select AM and DM data of oxygen and chlorophyll, and all DM data for nitrate. **AM data of nitrate are included after being corrected using**

CanyonB or WOA as explained in Johnson et al. 2021. For the three variables we use measures that are flagged as good, probably good, and interpolated.

comment3b

P8.l199: **"**when all four drift estimates agree in sign"

> Not clear what do you refer here by "four drift estimates". In P7.l196, it is mentioned that drift is evaluated at two different depths and what are the rest of two estimates? Or the number of four has nothing to do with that?

We chose two methods (RANSAC and Theil Sen) and two depths (600 and 800 m) to obtain a solid basis in evaluating if a float has an oxygen sensor drift, thus "four drift estimates" refer to the two methods applied at two depths. The calculation of the drift with two methods and at two depths avoids possible fake drift detections due to changes in the water masses. The clarification will be added in the new version of the manuscript (see **comment4b**).

comment4b

P8.l202-l203: "it can be assumed that O2 values at surface are already fixed by the GDACs"

> As is stated in p2.l41, not every O2 sensor is calibrated in air and air calibration is one of the most important calibration steps to make O2 data trustable. Do you believe O2 values at surface are fixed even for the old non air calibrated sensor data? Or are old sensor data not included in this specific study period, 2017-2018?

We used data coming from sensors calibrated both in air and in water and for those floats not calibrated in air an additional check on saturation is performed. When oxygen at the surface is far from the value of the oxygen saturation, profiles are excluded as explained in section 2.5. We would like to not add saturation check details in this paragraph that is focused on the novel Oxygen QC procedure. However, we propose to modify the paragraph as follows also taking into account other comments (**comment3b** and **comment15a**), new text in bold):

L. 195 Here, the optode sensor in situ drift is evaluated through non parametric methods (RANSAC and Theil Sen) at two different depths (600 and 800 meters) **to avoid possible fake drift detection because of changes in the water masses.**

[…]

L. 201 The presence of a drift is established when all four drift estimates **(RANSAC at 600 and 800 meters, Theil-Sen at 600 and 800 meters)** agree in sign and their average value is greater than 1 mmol m−3 y. This threshold is chosen on the basis of results in Bittig et al. (2018b). **Then, the drift is removed from the oxygen profiles by setting the computed drift average at 600 meters and linearly interpolating toward the surface, where drift is set equal to zero. According to Thierry and Bittig, 2021, there is a lack of specific tests at depth, while several (i.e., 14) tests applied near-surface are already performed by the GDACs. The presence of near-surface tests motivates our choice to reduce the effect of our correction at the surface.**

comment5b

P9.l223: “profiles can be excluded when model-observation misfit is higher than given thresholds”

> Does it mean some profiles are actually excluded during your DA run or this just describes online data selection system? I also assume “model-observation misfit” means innovation, is it right? Can you also justify the reason behind of this online data elimination procedure?

Yes, when innovation exceeds a threshold, we exclude the profile to avoid corrections that can introduce not stable dynamics. This has been discussed in Teruzzi et al., 2021. It is important to remember that this threshold is very permissive, thus very few data have been excluded. The excluded profiles for checks based on innovation ranged from 0.1% for chlorophyll to less than 1% for nitrate.

The text will include the suggestion to use “innovation” and will be changed as follows (in bold new text, as for comment17a):

L223 During the data assimilation, **profiles are excluded when innovation exceeds specific threshold rules.** For the chlorophyll the threshold is set to 2 mg m−3, for nitrate, the two thresholds are 2 and 3 mmol m−3 for 0-50 m and 250-600 m layers, respectively **(as in Teruzzi et al., 2021).** Oxygen thresholds are 30 and 50 mmol m−3 in the 0-150m and 150-600m layers respectively (**thresholds are roughly 3 times the standard deviation of the climatology computed on Emodnet data for the different sub-basins**). Exceeding values have to be found in at least 5 vertical levels in the specific layers. Exclusions are set to avoid corrections that can trigger unstable dynamics after the assimilation (Teruzzi et al., 2021, Storto et al., 2011, Sakov et al., 2017 and Waller et al , 2018). The excluded profiles ranged from 0.1% for chlorophyll to less than 1% for nitrate.

Additionally, we will change "misfit" with "innovation", introducing the term at old L137:

which relies on the innovation (i.e., the difference between the observations y and the model background xb)

comment6b

P10.l246-247: “While for the satellite comparison the model daily averages .. the model first guess is used for metrics based on BGC-Argo”

> I believe the choice of these different RMSE metrices between satellite OC data and Argo profiles are based on the experiment settings of Argo profiles being assimilated while satellite OC data are not assimilated. By choosing the first guess state to be compared with not yet assimilated Argo profiles, you can use the Argo profiles as independent data. If it were the case, better to describe so here.

Thank you for the comment. Since also the other Reviewers suggested (comment18a comment7b comment2c) to clarify this point, we propose the following text to include all the different comments:

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

comment7b

P10.l251: "Satellite L3 products from Copernicus Marine Service catalogue .."

> Usage of this data set requires proper citation. Plus, this sentence is floating without clear connection in 3.2. Does it mean satellite OC RMSE metric is based on this weekly averaged data? If so, does weekly cycle coincide with an analysis cycle? Please make its significance clear.

Many thanks for spotting this inconsistency in the text. In the present OSE experiment, we used daily L3 maps of satellite chlorophyll from the Copernicus repository for the model validation. The Copernicus OCEANCOLOUR_MED_BGC_L3_MY_009_143 product is given as daily maps, thus the comparison uses the model daily output. Please consider that we will change L246-247 lines also considering the comments of the other Reviewers as written in the comment above (**comment6b**).

comment8b

P11.l274-l275: "Here, improvements related to chlorophyll assimilation can be observed in nwm, ion and lev in winter and at depth in tyr, ion and lev in summer (Figure 5 middle panel)"

P11.l278: "the direct chlorophyll assimilation is more effective than .."

> Since there is no experiment with only assimilating chlorophyll in this study, it is not easy to point out degree of "improvements related to chlorophyll assimilation" and if direct chlorophyll assimilation is more effective than the dynamical model adjustment after nitrate assimilation. You need to provide extra analysis to support these statements.

Thank you for the comment. We do agree with the Reviewer that our text was in part speculative. The objective of the present work is to assess the impact of the addition of extra nitrate profiles from NN (DAnn). Thus, our present results show that there is no further improvement of RMSE of chlorophyll after DAnn with respect to the improvement shown by the DAfl run. The discussion about the relative impacts of nitrate vs chlorophyll profile assimilation have been already assessed in Cossarini et al., 2019 and Teruzzi et al., 2021. We will change the text of all the speculative conclusions. For example, we propose to change the text at L.274-280 as follows:

L.274-280

Since the DAfl and DAnn simulations share the same chlorophyll assimilative setup, the RMSE improvements with respect to the chlorophyll assimilation can be evaluated by comparing the HIND with DAfl or DAnn simulations (Figure 5 middle panel). For chlorophyll, a reduction of RMSE in DAfl (Figure 5 middle panel) is observed in nwm, ion and lev in winter and at depth in tyr, ion and lev in summer, which is in line to what already shown for nitrate and chlorophyll profiles assimilation in Teruzzi et al. (2021). The positive impact of DAnn on nitrate RMSE (Figure 5 top panel) is not directly transferred to the vertical chlorophyll RMSE (as observed for figure 4). This is due to the fact that direct chlorophyll assimilation is more effective than the dynamical model adjustment after nitrate and reconstructed nitrate assimilation in the areas close to the observed chlorophyll profiles.

comment9b

P13. "3.3.1 Impacts on biogeochemical vertical dynamics".

> There is no description on how figures 6, 7, 8 and 9 are plotted. Are they sub-basin averaged value? “the basin wide averages of DAnn display .. (Figure 6)” at P13.l310 infers these figures are basin-average, but it is never be stated clearly.

Yes, figures 6, 7, 8 and 9 show averaged values over two selected sub-basins (NMW and Ion2 in map of Figure 2) and the whole Mediterranean Sea. We will add a short explanation of Figure 6,7, 8 and 9 at old L287-289.
Following comments from other reviewers (in **comment20a**), the new version of the section reads as follows:

To assess the impact of profile assimilation in changing the vertical gradients of biogeochemical variables, the Figures 6, 7 and 8 show the Hovmoller diagrams of the spatial averages for two selected sub-basins (NMW and Ion2 in map of Figures 2) and for the whole Mediterranean Sea. This representation offers additional details on the vertical impact of the reconstructed nitrate profile assimilation with respect to the validation of Figure 5 that considers only model points corresponding to the location of BGC-Argo profiles. The two sub-basins represent two different trophic conditions in the Mediterranean Sea. The North Western Mediterranean (nwm) has higher level of nutrient concentrations and more intense surface blooms in winter (Siokou-Frangou et al., 2010, and Di Biagio et al., 2022). In summer, nwm has higher chlorophyll concentration at the deep chlorophyll maximum (DCM), shallow nitracline, and shallow subsurface oxygen maximum (SOM) (first column in Figures 6, 7, 8, and 9). On the contrary, more oligotrophic conditions and deeper nitracline and DCM are found in the eastern subbasin (ion2, second column of Figures 6, 7, 8, and 9).

Considering nitrate (Figure 6), the multivariate assimilation (DAfl) corrects a general positive bias of the model in all the Mediterranean areas (blue pattern in Figure 6). The addition of reconstructed profiles makes the corrections stronger. On average the nitrate concentration below the nitracline decreases by 8% and 11% in DAfl and DAnn runs, respectively. Both the assimilation runs also show changes of the nitracline depth (i.e., depth at which the vertical gradient of nitrate is maximum) with more intense deepening in the DAnn simulation.

Differences between the assimilation and the reference run accumulate over time. The rate of this accumulation is highest during the first year while during the second year it decreases and the differences remain almost constant in sub-basins with a high number of BGC-Argo and reconstructed profiles (e.g., NWM in Fig. 6).

On the other hand, considering other areas (e.g., ion) and the whole Mediterranean Sea, which comprises some under-sampled areas (e.g., southern Ionian and southern western basin), the effect of DA corrections is propagating after the two year(third column of Figure 6).

Very similar patterns are also observed in the Hovmoller diagrams of phosphate (Figure 7), which is an updated variable of the multivariate variational assimilation scheme through nitrate-phosphate covariance. In fact, the general negative corrections on phosphate fields are linked to the high positive values of the covariance matrix between nitrate and phosphate (Teruzzi et al., 2021).

Considering chlorophyll (Fig. 8), the main differences between DAfl and HIND are a slightly reduction of the DCM chlorophyll concentration (e.g., variation smaller than 5% with respect to HIND simulation) and a correction of the timing of the surface winter blooms (second row in Figure 8). Even if the chlorophyll validation (Figure 5) does not show significant differences between DAfl and DAnn, the basin wide averages of DAnn display more intense corrections with respect to DAfl in terms of DCM depth and chlorophyll intensity and overall chlorophyll concentration (Figure 8). Over the 0-200 m layer of the whole Mediterranean Sea, the chlorophyll decreases with respect to HIND are 4% and 5% for DAfl and DAnn, respectively.

Corrections on oxygen dynamics after the multivariate assimilation (DAfl, second row in Figure 9) are either positive or negative depending on the area and the period of the year. In particular, corrections are mostly positive in ion2, while the NWM sub-basin shows negative corrections in the subsurface layer and positive ones in the upper layer of the second year. On the Mediterranean basin-wide scale, the average correction is 0.2% for the 0-200m layer. The addition of the nitrate reconstructed profiles does not alter the correction pattern with an average correction of 0.3%. However, the largest differences between the two assimilation runs can be spotted in areas with a high density of reconstructed profiles during summer (e.g., NWM, first column in Figure 9). As observed in the nitrate and chlorophyll figures, the assimilation of reconstructed profiles causes a decrease of the summer productivity in the DCM layer. Consequently, less oxygen is produced generating the negative changes in the DCM layer in the bottom left panel of Figure 9. Because of the smaller amount of subsequent sinking organic matter, less oxygen is consumed in remineralization processes in layers below the DCM in late summer and autumn, and positive oxygen changes are generated, particularly during 2018.

P13.l297: “Nitracline depth”

> There is no definition of nitracline depth. Please be specific.

We will add the definition directly in the text. The definition of nitracline and phosphocholine will be added also in the captions of the Figure 7 and Figure 8.

At the Mediterranean scale, the nitrate concentration below the nitracline (i.e., the depth at which the nitrate is 2 mmol/m-3) decreases by 8% and 11% in DAfl and DAnn runs, respectively.

P13.l296-l297: “decreases by 8% and 11% in DAfl and DAnn runs, respectively”

> Contrary to the clear difference in impact of nitrate assimilation in DAfl and Dann at nwm in Figure 7, RMSE profiles in Figure 5 (especially Summer Nitrate at Nwm) does not show such difference in the two DA experiments. Can you explain why?

Figure 7 is plotted using the model average daily outputs computed over the whole area of the sub-basin, while Figure 5 reports the RMSE statistics computed on model background values of the grid points corresponding to the locations of BGC-Argo nitrate profiles. Moreover, consider that validation uses only BGC-Argo observations and not the reconstructed profiles (see text at lines 266-267). Figure 7 shows the impact of the extra nitrate profiles assimilation over areas that are not observed by the real BGC-Argo, which motivates the inclusion of the figures.

P13.l299: “eventually reach a stationary phase”

> What does it mean by a stationary phase and how do you measure it?

Indeed, the term stationary was misleading. We refer to the fact that during the second year, the rate of accumulation of DA impact decreases or is quite null. The two-year simulation shows that the assimilation corrects the initial model bias in most of the areas (e.g., the most observed sub-basins such as ion2) but that other undersampled areas require more time to be influenced by the assimilation. We propose the following modified version of the sentence (see also **comment9b**):

Differences between the assimilation and the reference run accumulate over time. **The rate of this accumulation is highest during the first year while during the second year it decreases and the differences remain almost constant in sub-basins with a high number of BGC-Argo and reconstructed profiles (e.g., NWM in Fig. 6).**

P13.l306: “As a consequence of both the direct assimilation of chlorophyll profiles and the dynamical model adjustment after nitrate assimilation”

> Again, how can you argue a consequence of dynamical model adjustment only from nitrate assimilation with this OSE settings? For example, why would phytoplankton biomass change as a consequence of direct assimilation of chlorophyll not affect chlorophyll concentration in the DCM as a consequence of its dynamical model adjustment? If extra material not provided, this statement is speculative.

Thank you for the comment. The second Hovmoeller of Figure 6 shows the difference between DAfl and HIND runs. This difference is due to the direct impact of chlorophyll assimilation and indirect effects (e.g., model dynamical adjustments) after nitrate assimilation. It is correct that we can not distinguish between the two runs (DAfl and DAnn), however the original sentence was meant to highlight the difference between DAfl and HIND. To avoid any misinterpretation, we will change the sentence in L306 as follows (see also **comment9b**):

Considering chlorophyll (Figure 8), the main differences between DAfl and HIND are a slightly reduction of the DCM chlorophyll concentration (e.g., variation smaller than 5% with respect to HIND simulation) and a correction of the timing of the surface winter blooms (second row in Figure 8).

P13.l313: “oxygen profiles assimilation (DAfl, second row in Figure 9) provides positive or negative corrections”

> As is described by authors in the subsequent sentences, changes in phytoplankton biomass also change oxygen through primary production and remineralization process as dynamical model adjustment. Thus, assimilation of chlorophyll and nitrate both have a potential to alter oxygen. How can you judge what can be seen in Figure 9 is sole consequence of oxygen assimilation? This sentence contradicts with statements following about impact if reconstructed nitrate profile assimilation in oxygen.

We do agree with the reviewer. By comparing DAfl and HIND we can only provide an assessment of the overall impact on oxygen of the multivariate (nitrate, chlorophyll and oxygen). We will change the sentence at LL313-314 as follows(see also **comment9b**):

Corrections on oxygen dynamics after the multivariate assimilation (DAfl, second row in Figure 9) are either positive or negative depending on the area and the period of the year. In particular, corrections are mostly positive in ion2, while the NWM sub-basin shows negative corrections in the subsurface layer and positive ones in the upper layer of the second year.

P13.l316-l318: “The only noticeable difference .. > This is one of the most important findings in this study as an impact of reconstructed Nitrate profile assimilation, but difference between DAfl and DAnn in figure 9 (summer period in NWM) can not be found in RMSE profiles in figure 5 (summer

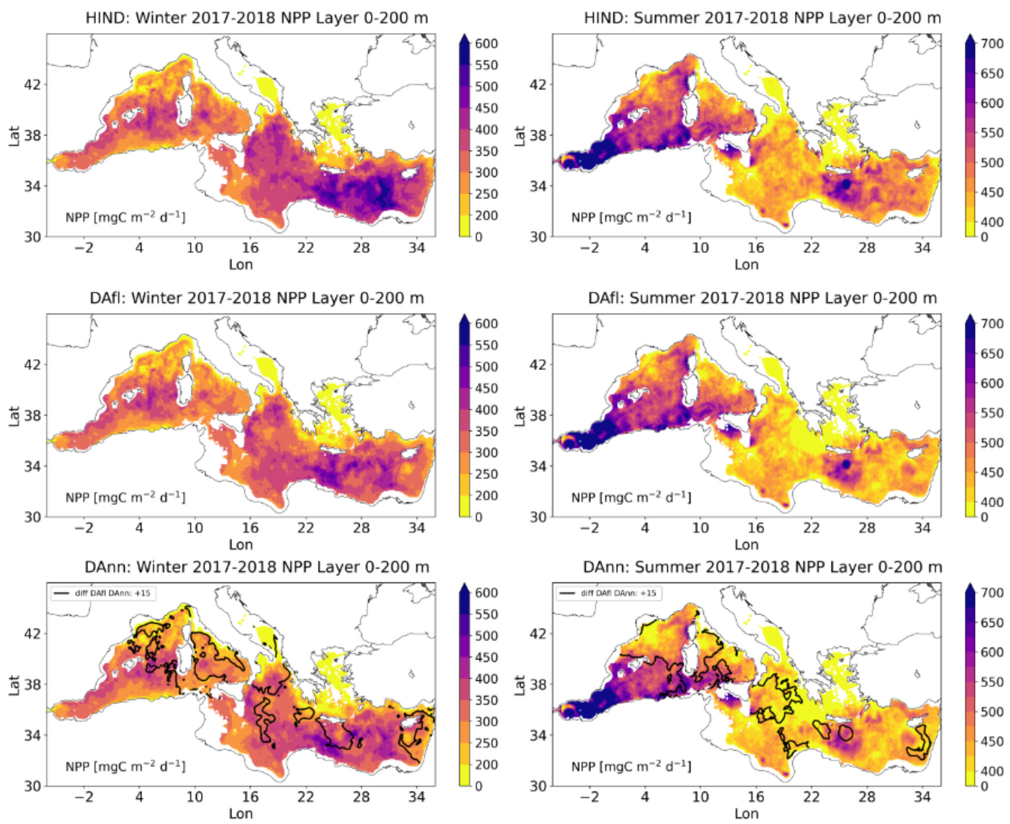
Oxygen in Nwm) and we can not judge if this difference in DAnn against DAfl is improvement or not. Can you explain why?
Thank you for the comment. Figure 9 is plotted using the model's average daily outputs computed over the whole area of a given sub-basin, while Figure 5 reports the RMSE statistics computed on model background values of the grid points corresponding to the locations of BGC-Argo oxygen profiles. Given that oxygen profiles are assimilated in DAfl, the positive message is that the assimilation of extra nitrate profiles does not degrade the quality of DAfl (with oxygen assimilation). It is reasonable to expect that DAnn can perform at maximum as good as DAfl with respect to oxygen. On the other hand, Figure 9 shows an additional detail: nitrate extra profile assimilation can provide changes to the model dynamics that is wider than oxygen one. This change can impact oxygen in areas distant from the location of oxygen BGC-Argo profiles. The text is changed following all reviewers' comments (see also **comment9b**).

comment10b

P16 entire section of 3.3.2

> Since difference between DAfl and DAnn is almost impossible to see in Figure 10, readers can not confirm what is described in this subsection. Please reevaluate how to present different impact of DA settings in NPP.

Following the suggestions of the Reviewers (see **comment21a**) and in order to better highlight the NPP differences between the three simulations, we changed the colorbar and added the15 mg/m3 contour line on the last row. The contour refers to the difference between NPP in DAfl and NPP in DAnn.



comment11b

P16.l339-l341: “In fact … after chlorophyll assimilation”

How can you measure that weak negative correction of macronutrients is the main cause of reduced NPP outweighing the effect due to change in phytoplankton biomass after chlorophyll assimilation? As far as I read, there is no concrete material supporting this statement is provided. Unless extra material provided, this statement is speculative.

Thanks for the suggestion. We will remove the speculative conclusion on chlorophyll assimilation in L 339-341.

The weaker fertilization of the surface layer in DAnn, which occurs for both macronutrients after assimilation (Figures 7 and 8), causes a reduction of the net primary production.

comment12b

P17. 3.3.3

> In figure 11, figure title indicates Nitrate Iij(t) is evaluated over 0-600m depth range rather than 0-300m depth specified in equation (2). If it were the case, please specify so. If not, please fix the figure titles in figure 11.

Thank you for spotting the typo. We used 0-300 and 0-600 m layers for chlorophyll and nitrate respectively. We will correct the text and the equation title of the figures. We will replace subscript “300” with a more general term “maxdepth” in the equation 2.

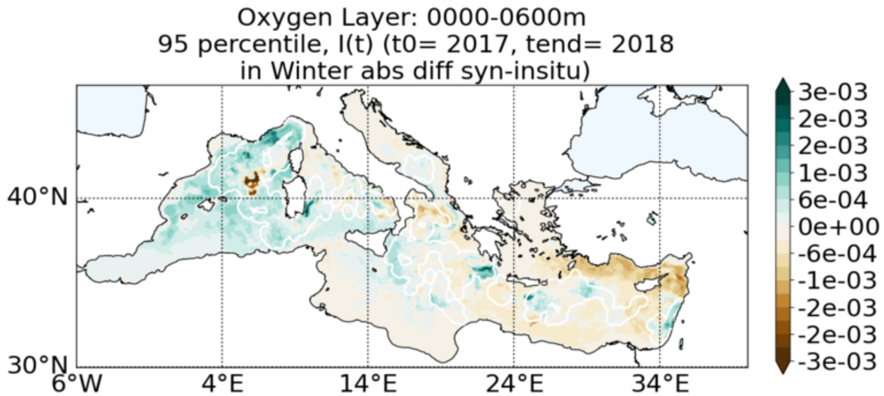
In this work, we adopt the impact indicator Iij (t) described in Teruzzi et al. (2021). The impact indicator allows quantifying the integrated response of assimilating BGC Argo profiles with respect to the no assimilative run: [equation_corrected..]. HIND is the reference, while Sim refers to one of the different DA setups (DAfl or DAnn). |Simij (t) − HINDij (t)| is the absolute difference between simulations (for each day and grid point), **while the subscript maxdepth represents the integral over the 0-300 m and 0-600m for chlorophyll and nitrate, respectively.**

comment13b

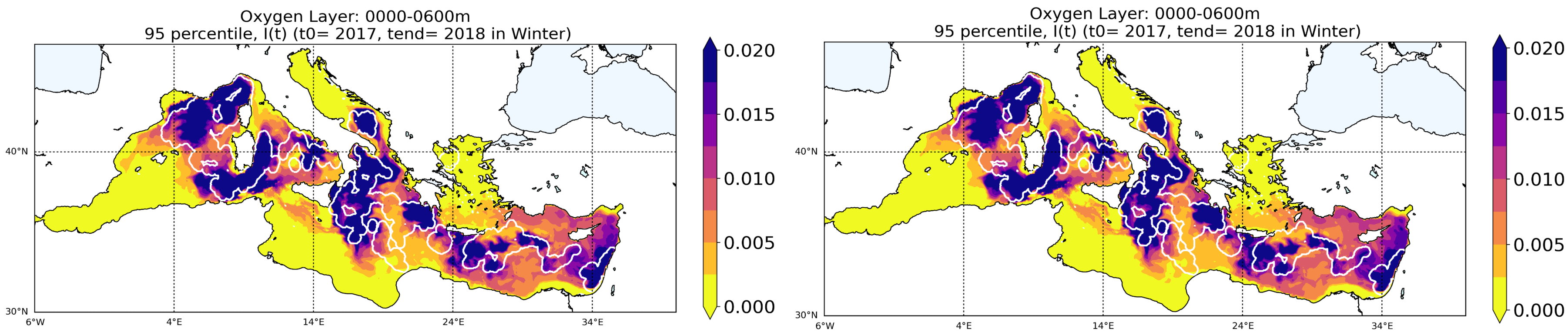
P17.l365: “since the same QC oxygen dataset was assimilated in DAfl and DAnn”

> But the authors just described in P13.l316-l318 that impact of the reconstructed Nitrate is noticeable in oxygen at least at NWM where density of the reconstructed Nitrate is large. Then it does not make sense that you do not see difference in the two DA experiments. Why do you not see the difference in the Iij 95th percentiles maps for oxygen?

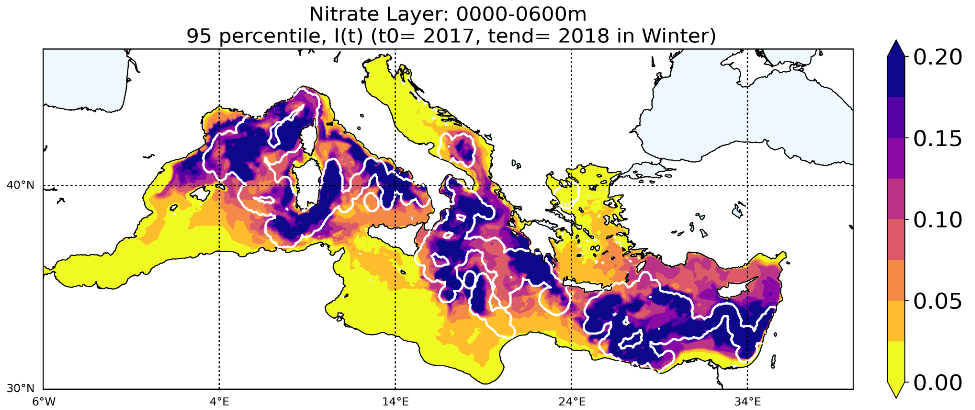
Thank you for your comment. We observe low differences (10^{-3}) mainly in the NWM, between the two DA experiments (DAfl and DAnn) maps. Since the “*oxygen assimilation updates only the oxygen itself*” *(L 154 and* **comment11a**) these differences are due to the reconstructed profiles of nitrate that cause a decrease of productivity, a loss of oxygen production and a loss of remineralization. In the new version of the manuscript, we will elaborate on this topic, also linking to **comment13a** (the spatio-temporal availability of nitrate and reconstructed nitrate). For sake of clarity, we also provide hereafter the differences between the Impact indicator for DAfl and DAnn of oxygen:



Moreover, as commented in **comment11a,** in our DA scheme, the oxygen dynamics is not directly affected by chlorophyll and nitrate increments by the assimilation. This explains how low the differences between the DAfl and DAnn oxygen impact maps are (Winter Impact map DAfl left-DANN right):



As written in the text L364-365 the *oxygen impact maps (not shown) are very similar to nitrate DAnn maps:*



PP19.l396-l.397: “In this work, important impacts are also observed in summer for all variables, as a consequence of the increased number of assimilated profiles.”

> It is not clear what does it mean by “a consequence of the increased number of assimilated profiles”. Increased number of nitrate profile from DAfl to DAnn? Or about something else? As far as I understand, main reason why we see impact of DA in summer in DA experiments in this study compared to Teruzzi et al. (2021) is because satellite OC can not see DCM while Argo float profiles see the signal by multiple sensors. In that sense, you could see the impact of Argo profile assimilation no matter how small or large number of profiles is. Please reevaluate this statement.

Thanks for your comments. By comparing the maps of impact indicators for DAfl and DAnn, we show the potential additional benefit of extra nitrate profiles in a multiplatform data assimilation simulation. We will better explain this concept as follows:

In Teruzzi et al. (2021), results of the impact indicator principally showed the efficiency of ocean color assimilation in constraining chlorophyll dynamics especially during winter and the benefit of profile assimilation during summer. In this work, we show that the potential benefit of profile assimilation in summer would be bigger and wider because of the assimilation of the additional NN-reconstructed nitrate profiles.

comment14b

PP19.l399-l400: “Indeed .. box every 10 days”

> I do not understand which “results” in this study support this statement. Basin coverage rate of BGC-Argo floats equipped with oxygen sensors is simply determined by deployment plan. Or do you like to say that the new O2 QC module prove enough number of O2 profile survives to be ingested to nn module? I read 3.1, but could not get such information. Please be clearer about meaning of this statement.

Our aim here is to highlight that the OSE experiment shows that the basin coverage rate of nitrate can potentially be as high as the BGC-Argo equipped with an oxygen sensor. Considering also comments of the other Reviewer (see also **comment23a**), we will better explain this concept at L399-400:

Through the integration of NN and DA, the number of nitrate profiles ingested can potentially be as high as the BGC-Argo equipped with an oxygen sensor (i.e., more than double of the nitrate profiles), which corresponds to a density of 1 profile in each 2.5deg x 2.5deg box every 10 days for the 2017-2018 period.

comment15b

PP19.l401-l406: "while, up to … by a 3D varying correlation radius (Storto et al., 2014)">

This discussion on improvement in meso-scale dynamics look out of topic and I can not see the reason why it is needed to be discussed here. Especially confusing knowing that 2.5 degree by 2.5 degree horizontal resolution in BGC profiles potentially could be achieved by nn with oxygen profile is far below meso-scale resolving resolution of o (50km).

Thank you for the comment. By redefining horizontal covariance error we can only increase the spatial area in which each float has an impact. We will rephrase as follows considering all the reviews (in **comment24a**):

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

PP19.l418: "0.50 mmol2 m−3 for nitrate"

> This information should be included in 2.2.

OK, we will add the information.

PP19.l423-l.429: "Indeed … Li et al.(2021)»

MLP base Sauzède et al. (2017) overcame of this issue by adding pressure as input variables in MLP. Why do you believe choosing other NN approach such as 1D CNN is important before using pressure or depth information in MLP-NN-MED?

As shown in Pietropolli et al., 2023 (GMDhttps://doi.org/10.5194/egusphere-2023-1876), MLP does not explicitly consider that close points in a profile share information (the back propagation during the training treats two close points in a profile as not-correlated values of the target variable). As a result, a profile reconstructed with MLP and T, S, O2 and pressure input from BGC-Argo can show discontinuities that need to be filtered with additional steps in the procedure (see line L176-179). This potential pitfall is overcome by 1D convolutional NN, which learns explicitly the shape of the vertical profiles during the training, thus exploiting the fact that each point of a profile shares information with its neighbors.

In Pietropolli et al., 2023 there is also a comparison between vertical profiles predicted through MLP-NN-MED and PPCon, which is the proposed 1D

CNN approach. Results demonstrate that changing the architecture leads to more smooth profils, which better approximate the original sampled vertical profiles.

Technical corrections:

P1.l20: “The Array for Real-time Geostrophic Oceanography”

> Please do not use this acronym for Argo. It is not official. There is a historical background why it should never be and I have it on the authority of one of the program founders who was present on the day the Argo project was first conceived: “Argo was named as a companion project to the proposed Jason altimetric satellite missions. The words indicating a putative interpretation of the letters Argo, Array for Real-time Geostrophic Oceanography, were created in a jocose moment while celebrating in a bar afterwards. It would be best to let an idea die whose origin was mediated entirely through the action of alcohol. Argo is not and was never meant to be an acronym. It should be written "Argo" and never as "ARGO".

Thanks for the clarification. We will remove it.

P1.l2: “and successfully integrated in”

> “and are successfully integrated in”

Ok, We will correct the sentence.

P7.l182: 2.4 BGC-Argo data and post-deployment oxygen quality control

> I assume subsection 2.4 is about QC-O2 module, but the module name is never referred to in this section but found in the next section, 2.5. Please make it clear that this is about QC-O2.

We will correct this inconsistency.

P7.l184: “and DM data were selected for oxygen and chlorophyll, while exclusively DM data were considered”

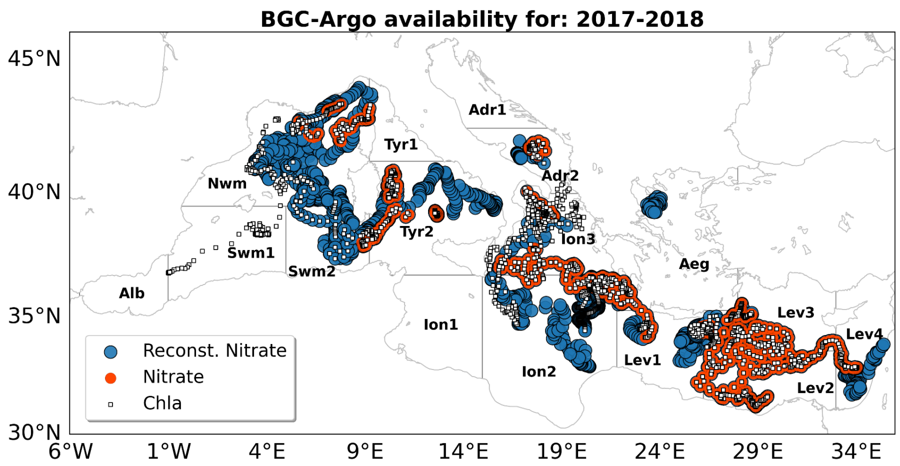
> Use only one expression among “delayed mode” or “DM”. Not together in this sentence, but the same unification of usage of acronym would be better for sets of “Adjusted/AM” and “Real Time/RT” for the entire this manuscript after RT, AM and DM defined at p2.l32.

Thanks for the comment. We will use the full name the first time and after the acronym.

P8.Figure 2: Coordinate labels font is too small and almost unreadable. Please enlarge its size.

> It is almost impossible to distinguish the three dots in the figure. Please consider using different colors or separate maps for each type of profile.

We propose the following new figure, with new colors, new markerstyles, edges and font-size. We have tested the effectiveness of the colors at: https://www.color-blindness.com/coblis-color-blindness-simulator/.



P8.Figure 2: "lev=lev1+lev2+lev3+lev4; ion=ion1+ion2+ion3; tyr=tyr1+tyr2; adr=adr1+adr2; swm=swm1+swm2"

> As far as I can read, this is the only place where aggregated sub- or macro- basins are defined. This should be properly defined in a table as suggested below. It is also recommended to use either "sub-" or "macro-" for minimizing confusion.

We will add the definition of the basins in the text.

comment16b

P9.l217-l220. "Finally, oxygen …"

> I can guess, this long sentence is hard to understand. Needs reorganization with shorter sentences.

We propose the following rephrased version of the paragraph (as for **comment16a**). In bold new text:

Before integrating data in the 3D-VarBio, the same pre-assimilation assessment described in Teruzzi et al. (2021) is applied to the chlorophyll profiles. **Nitrate profiles are rejected if concentration at the surface is higher than 3 mmol m−3.**

At surface, the oxygen profile exclusion is evaluated by calculating the difference between the uppermost oxygen measurement and the oxygen saturation (derived from temperature and salinity data from the Argo dataset as in Garcia and Gordon 1992). Profiles are excluded when this difference reaches the threshold of 10 mmol m−3.

At 600 meters, the difference between oxygen and a climatological reference oxygen at depth is calculated. Profiles are excluded when the difference reaches the threshold of 2 times the standard deviation of the same reference dataset. As reference dataset, we chose the EMODnet2018_int data collection that integrates the in situ aggregated EMODnet data (Buga et al., 2018) and the datasets listed in Lazzari et al.

(2016) and Cossarini et al. (2015b). The EMODnet2018_int dataset is available for 16 sub-basins (see Figure 2) in the Mediterranean Sea.

P9.l229-l.230: “The oxygen post-deployment quality check method” and “The post deployment oxygen QC method”

> I assume again, the QC method is referring QC O2 module described in 2.4 or not? If it were the case, please specify so explicitly. Plus, two different ways to refer the QC O2 module at the title of 3.1 and body of 3.1 is strange.

We will change the title of the Section and clarify this aspect in the introduction.

P10.l248: “is evaluated in winter (from February to April, FMA) and summer (from June to August, JJA)”

> Since your experiment period is two years from Jan 2017 to Dec 2018, do you use both 2017 and 2018 results for this evaluation?

Yes, we use both 2017 and 2018 results. Following all the reviewers' comments, we will revise this paragraph and add this information as well.

P10.l255. “the eastern sub-basins”

> Please define which sub-basins (lev1, lev2,…etc) are included in the definition of the eastern sub-basins.

P8.Figure 2 caption: “lev=lev1+lev2+lev3+lev4; ion=ion1+ion2+ion3; tyr=tyr1+tyr2; adr=adr1+adr2; swm=swm1+swm2”
P10.l257: “alb, swm and nwm”
P11.l263: ” Alboran, South West Mediterranean, North West Mediterranean, Tyrrhenian, Ionian and Levantine Seas”
P11.l271: ”is observed in nwm and tyr (winter) and in ion (summer).”
P11.l275: “in nwm, ion and lev in winter and at depth in tyr, ion and lev in summer”

> Association of long and short names of each sub-basin such as Alboran (alb), South West Mediterranean (swm) etc. is never clearly defined in this article. Please do in section 2.3 or add extra table to do so.

We will add all the information in the text. In particular we will explaining the sub-basin, the aggregated-basin and East-West basins organization.

P11.Figue 4.

> Figures are too small that it is hard to distinguish three bars at each domain. Please use larger size of figures.

We will enlarge the figure.

P12.Figure 5.

> Figures are too small that it is hard to distinguish three profiles especially between DAfl and DAnn. Plus, many figures do not have x axis labels. Please use larger size of figures or reconsider different way of presentation such as scatter plots and tables of RMSEs at selected depths.

Unfortunately, the lack of the x-axisis is due to a Latex-formatting error in the figure. We will increase the space between the rows to avoid this problem. We will also consider how to improve the readability of this figure.

P13.l288: “two sub-basins”
 > Please specify names of “two sub-basins” here before referring Figure 2.
We will specify the “two sub-basins” nwm and ion2.

Figure 6, 7, 8, 9
 > Figures and font sizes are too small. Be more specific about definitions of the second row and the third row in figure caption.
Taking into account the comments from all the reviewers, we will redo the figures and change the order of them (i.e., nitrate, phosphate, chlorophyll and oxygen). We will markedly revise section 3.3.1 to improve its readability. New version of the section is proposed at: **comment9b**.

P16. Figure 10. Name of experiments, HIND, DAfi and DAnn are Hind, DaIns and Dasyn in y axis label in the figures. It is confusing.
We will correct the typos in Figure 10.
P17.l348: “Here, HIND is here the reference”
 > “Here, HIND is the reference”
We will correct it.

P19.l367-l.389: Five paragraphs about oxygen QC.
> This information do not fit to “Discussion”, but rather should be integrated to 2.4.
We will remove part of these paragraphs (lines 382-386) since it is already presented in section 2.4. However, we think that it is relevant to highlight that the proposed drift assessment does not interfere with multidecadal shifts in the Mediterranean. Indeed, this work on the O2 QC procedure has been presented at several conferences (e.g., ADMT meeting in Hobart 24th of Octubre 2023) and the debate on how to correct oxygen sensor drift is still open. Finally, we would like to keep in the discussion a paragraph concerning one possible further development of the O2 QC procedure (i.e., oxygen drift analysis applied to fixed isopycnals).

P20.l408: “BGC-Argo OS”
 > Please define meaning of OS. Observing system?
We will correct the acronym with the extended name *Observing system*.