#REV 3 Julien Brajard, 10 Oct 2023

The work presented in this paper is of interest to the community, as outlined by the two Reviewers. Nevertheless, as it is noted by the reviewers, the paper needs to be clarified, and the main message more clearly conveyed. I hope that all the comments and suggestions by the reviewers will help to provide an improved revised version.

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**).

Other comments:

comment1c

Section 2.3 I agree with Reviewer 1 that details about the neural net approach are missing. Especially the sentence "incorporating nonlinear functions, adjusting neuron count, and optimizing the training algorithm" needs to be expanded, since we could wrongly understand that the Fourier et al. approach does not incorporate nonlinear functions (while in reality, they use the nonlinear sigmoid function).

Ok, we have revised the entire section as follows (also considering suggestions of the other reviewers, e.g., **comment12a**):

The NN-MLP-MED (Pietropolli et al., 2023) is the evolution of previous MLP architectures developed to predict low-sampled variables (e.g., nutrients) starting from high-sampled ones (e.g., temperature) (Sauzède et al. 2017, Bittig et al. 2018c, and Fourrier et al. 2020). NN-MLP-MED is a deterministic Feed-Forward Neural Network based on a MLP structure. The NN-MLP-MED consists of the merging of 10 different MLP architectures, each one with the same input and output features, composed by the same number of hidden layers (i.e., 2), but composed by a different number of neurons per layer. The final prediction resulting from the NN-MLP-MED is the mean of all the predictions of these components.

The data flow of this MLP-based approach follows the forward direction from the input to the output layers through the neurons which composed the layers.

In our OSE experiment, the trained NN-MLP-MED reconstructs nitrate profiles (output) from temperature and salinity (Argo), oxygen (BGC-Argo) and float date, latitude and longitude (inputs).

The NN-MLP-MED model presents some novel elements with respect to the mentioned methods (and in particular with respect to Canyon-Med in Fourrier et al. 2020). Firstly, the input dataset includes a larger sample size and wider coverage of the Mediterranean Sea region, i.e., the quality-controlled EMODnet2018_int data collection which integrates the in situ aggregated EMODnet data (Buga et al., 2018) and direct observations (i.e., campaigns) as in Lazzari et al. (2016) and Cossarini et al. (2015b).

Secondly, the quality of the input dataset benefits from a two-step quality check process, removing noisy and unreliable samples. The neural network architecture was also modified to enhance prediction performance by accurately selecting a performing nonlinear function, adjusting and optimizing the amount of neurons for each layer of the MLP model, and choosing a different optimization strategy to train the algorithm. NN-MLP-MED also includes a vertical smoothing (running mean of 5-10 m window) step and a climatological adjustment at depth 600m that is derived from EMODnet (Salon et al., 2019).
The input nitrate dataset for assimilation contains 938 BGC-Argo profiles and 2146 reconstructed nitrate profiles (Table 1). The reconstructed nitrate profiles are located 61% in the western and 39% in the eastern Mediterranean Sea, thus providing a larger and more homogeneous spatial coverage asin Figures 2.
Uncertainty of reconstructed nitrate associated to the EMODnet validation dataset is 0.5 mmol m−3, while it reaches 0.87 mmol m−3 when it predicts the BGC-Argo dataset (Pietropolli et al., 2023).

L246 "the model first guess" does it correspond to the background?
Yes, the first guess is the background. It is the state of the system before the assimilation. Given that BGC-Argo floats have a profiling (or measurement) frequency of nearly 10 days, the first guess corresponds to the 10-day predictions in the local areas around the location of a given profile. Considering also other comments, the text at old L246 will be changed as follows:
"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. "

About the assimilation: how frequent is the assimilation update? Is it 10 days?
Generally, it is 5 days. A very small number of floats and for a limited period of time (less than 20%) can have higher frequency based on the decision of PIs of the single floats. We will improve sentence at L189-191 including temporal frequency information of floats as follows:
All the three BGC variables have a fairly homogeneous spatial coverage between the western and eastern Mediterranean Sea, except for few areas not covered (see Figure 2) and a generally 5-day temporal sampling frequency. Higher sampling frequencies (< 5 days) are registered for the 20% of profiles, while <10% have a daily frequency.

comment2c
About the validation: Can you comment a bit on the choice of using the RMSE between BGC-Argo profile and model first guess as a validation. Since a previous measurement of a BGC-Argo profile was already assimilated, can a new measurement be considered independent? It could be interesting to have a quick discussion about the lagrangian autocorrelation…
Thanks for the comment. Due to the lack of independent in situ data, our validation has used the common practice of comparing the first guess with assimilated observations (Hollingsworth, et al., 1986). Additionally, the BGC-Argo floats generally have a frequency of 10 days, which makes the use of the first guess an evaluation of a 10-day prediction in the local areas around the location of a given profile. Temporal autocorrelation between subsequent profiles for the same floats was evaluated for chlorophyll in Cossarini et al., 2019. Those results (Figure 11 in Cossarini et al., 2019)

showed that the persistence of the corrections has a half decay of about 4–5 days. Given some computation limitations this metric has not been tested for O2 and NO3. We will comment this point at L246-247 as follows (also considering suggestions at **comment18a** and **comment6b** **comment7b**):

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. An analysis of the persistence of the corrections of chlorophyll showed that the half decay is of about 4–5 days (Cossarini et al., 2019), which makes the frequency float sampling (e.g. 10 days) big enough to reduce the risk of autocorrelation.