PPCon 1.0: Biogeochemical Argo Profile Prediction with 1D Convolutional Networks

Response to the Editor

Dear Editor and Reviewers,
We appreciate the constructive comments and suggestions provided by the reviewer. Following their suggestions, we present our point-by-point responses to the reviewer's comments.
The reviewer's comments are highlighted in blue, followed by our responses in black. In each response, we detail the proposed changes to the manuscript, including any modified text and/or figures (in red).

Dear Gloria Pietropolli, Luca Manzoni, and Gianpiero Cossarini,

I would like to thank the authors for their efforts to improve the manuscript. It now reads much more smoothly and the consistent use of terminology to describe items/aspects makes it much easier to understand.

The Med Sea focus is still a bit hidden, only to appear in the last sentence of the abstract. I would prefer this info to be presented more up-front, but I guess this is at the author's discretion.

(R1) However, there is one aspect where I dissent with what the authors claim. They write: l. 8f.: "However, MLPs lack awareness of the typical shape of biogeochemical variable profiles they aim to infer" ('claim a') and continue: l. 9f.: "resulting in irregularities such as jumps and gaps when used for the prediction of vertical profiles." ('claim b')

(**) We thank the reviewer for these observations, we will discuss them further and make the appropriate changes in the text.

Nevertheless, before proceeding, we would like to point out that the goal and key messages of our paper target three variables: nitrate, chlorophyll and bbp700   The real strength of convolutional architectures (i.e., explicitly considering 1D vectors) compared to MLP architectures becomes evident at least for the last two, where convolutional architectures can learn the shape of the vertical profile, making it possible to predict more vertically-dynamic variables like chlorophyll and bbp700..

Regarding the nitrate variable, based on our experience with training and testing MLPs and CNNs, we noted that CNNs can be a viable alternative to previously proposed MLP approaches. These considerations include both the performance obtained with CNNs and the less demanding process of creating the architecture. Creating an MLP architecture requires more oceanographic domain knowledge (e.g., the preprocessing in CANYON-MED, where each input needs specific functions) and a longer hyperparameter tuning phase, while CNNs are naturally suited for this kind of application, making the tuning phase less demanding.

Said that, we agree with the reviewer that MLP approaches for nitrate have shown very good performance, so the effectiveness of CNN and MLP approaches is similar.

Regarding the two claims:

Claim a: MLPs, from a theoretical point of view, operate point-wise. However, we agree that if the input changes smoothly, MLPs will provide a smooth output. On the other hand, 1D-CNNs, by construction, can deal with profiles. Our objective is to demonstrate that 1D-CNNs can work as well as previously proposed approaches (again, when referring to the nitrate variables). We will rephrase the sentence in the abstract as follows: "Although MLPs can produce smooth outputs if the inputs change smoothly, 1D-CNNs are inherently designed to handle profile data effectively."

Claim b: The sentence applies not only to nitrate but to all variables from BGC-Argo. Chlorophyll and bbp700 can benefit significantly from a 1D-CNN approach, avoiding undesired jumps in profile reconstruction (see Figures 4 and 5 and figures in Pietropolli et al., 2023a). However, we agree with the reviewer that for nitrate, MLPs perform as smoothly as 1D-CNNs. We will change the sentence in the abstract (lines 8-10 of the previous version) as reported in the previous point.

As written in my previous comment, (1.) I don't think either of the two statements is true, and (2.) I don't see evidence presented by the authors to convince me otherwise (more on 2. later).

I see and appreciate in the remainder of the manuscript that these two claims have been toned down, compared to the initial version. However, I don't think they are justifiable.

To claim a: The authors are correct in that MLPs act and are trained on point-wise data, whereas CNNs take strides of data (e.g., profiles) and consider both their value and their arrangement (e.g., profile shape). To use CNNs to predict some profile data from other profile data, taking benefit of the shape of that other profile data, is a (promising) step forward compared to MLPs predicting some profile data. However, that does not mean that MLPs for Ocean prediction are agnostic to their neighbouring data points. Instead of an explicit shape awareness like CNNs, MLPs have an implicit awareness of the typical shape of profiles they want to infer. Why? Because the Ocean is smooth.

The parameter space in the Ocean is continuous and smooth (with a large thanks to mixing). Just for illustration: Below two T-S diagrams for two of the floats used. (Quality-controlled) Ocean data are smooth to start with. Going along a profile step-by-step (both in parameter space or against depth) gives only small, step-wise modifications of the variable to be predicted and the variables used as predictors alike. Which means that, at any given point in parameter space, you have a certain idea of how your environment will look like: Probably not an awful lot different, but just a little. I.e., even if only given point-wise knowledge at a time, MLPs do have an awareness of how their profile will look like nearby (i.e., not a lot different). I.e., there is (some) spatial awareness of profile shape in MLPs, too, due to point-wise proximity in parameter space and due to the Ocean's parameter space being smooth.

To claim b: Starting off from (i) a smooth parameter space (previous point), and (ii) using a neural network (MLP) architecture complexity suited for the size of the training data as well as (iii) ensuring that there is no overfitting by properly selected regularization, then the (MLP) neural network outcome must be an approximation of the parameter space trained on. If properly regularized, the neural network is by definition smoother than the training data from (i). If not overly complex, then the network's weights and parameters are sufficiently constrained by the training data from (i), so that the (MLP) network represents a continuous function (no poles, irregularities, gaps). If the training data are smooth to start with, then that cannot cause irregularities or gaps either.

So I am left with claims where my arguments run against them. The two claims are brought up again in l. 68-73 but without illustration. So when coming to Figure 3 (nitrate profiles for selected floats), I wondered why the measurements and PPCon are shown but the Pietropolli et al. 2023a MLP comparison (as well as other MLPs for comparison like CANYON-MED, CANYON-B) were dropped?

We decided to present the comparison with other methods in Appendix B for several reasons. The most important reason is that this paper aims to predict three variables: nitrate, chlorophyll, and bbp700. The comparison with other methods is only possible for the nitrate variable, as there are no MLP architectures capable of adequately predicting the other two variables. Therefore, since nitrate is the only variable that allows for a meaningful comparison, we chose to include this information in the Appendix.

Thank you for adding the WMO and profile numbers, which allowed me to go search for the data and do those comparison plots on my side. Here's what I get for the measurements as well as the two MLPs where I had access to the code for prediction: CANYON-MED (specifically trained on the Med Sea with a dataset extended beyond GLODAPv2) and CANYON-B (trained with very little Med Sea data from GLODAPv2, so no great performance to expect; but it provides confidence intervals, which I find instructive):

Same as Fig. 3 but with measured float data directly from the GDAC and two MLPs added (dark blue: CANYON-MED, light blue: CANYON-B).

From the comparison, I find it rather comforting and confirming my line of argument that the MLPs give a similarly smooth profile shape as the CNN (in the manuscript Fig. 3) and no irregularities or jumps in the profile – unless governed by the water mass properties (e.g., middle profile just shallower than 50 dbar) and seen in the measured nitrate, too. (As said, I could not confirm/falsify whether the EMODNET-based Pietropolli et al. 2023a MLP shows irregularities or jumps in the profile.)

(R2) What I find a bit discomforting is that the measured data in my case looks not the same as shown in Figure 3 of the manuscript: All profiles in Fig. 3 are much smoother (e.g., base of the mixed layer is significantly eroded; interleaved water mass around ~50 dbar in middle panel fully absent) and also the bottom portion of the middle profile has a different shape (and value). I have confirmed the profile's locations and dates that I used – they match to the floats and cycle numbers (Thank you for Table 5)!

Thank you for pointing out that the bottom portion of the middle profile of Figure 3 has a different shape and value. After double-checking, we noticed a typo in the profile name (i.e., 6901769 instead of 6901768). We have updated Figure 3 by changing the float name to the correct one (6901769_083).

(*) Regarding the second concern about the smoother profiles in Figure 3, this is due to the discretization performed on these vertical profiles. Specifically, for the nitrate variable, we considered profiles ranging from 0 to 1000 meters, and for the other two variables, we considered profiles ranging from 0 to 200 meters. We used equal discretization for the input and output which was necessary for the functionality of the convolutional architecture. Thus, a preliminary task of interpolation from the irregular and varying profile depths of original data to a regular discretization was performed.

We discretized the vertical profiles into 200 points, resulting in a 1-meter resolution for chlorophyll and bbp700 and a 5-meter resolution for nitrate. The 5-meter resolution, especially in the upper part, led to a smoother appearance compared to the original vertical profile, where the sampling was higher than once every 5 meters.

The discretization and interpolation are well explained in lines 222-225 in the "experimental setting" section and all reconstructed profiles (either MLP and CNN) in Appendix B have been produced with the same discretization to allow a fair comparison.

We will better clarify this point by adding a sentence in Appendix B (old lines 405-407) as follows:

The three ML architectures are: the 1D CNN of the present work (PPCon), MLP trained on point-wise data from Emodnet (Pietropolli et al., 2023a) and MLP trained on point-wise data (Fourrier et al., 2020). Input data from Argo and BGC-Argo for all approaches have been interpolated to the regular 5-meter discretization as explained in the section 4.2 Experimental Settings.

The two claims are then taken up again in the results (l. 317-318; no illustration) and later on in the discussion (l. 411-412; no illustration) and the reader is referred to Appendix B (l. 392; l. 414).

This Appendix B (comparison between reconstructions by PPCon and MLP architectures) is very promising. And after experience with Fig. 3 I tried to redo the figure B1 on my end, too:

Same as Fig. B1 but with measured float data directly from the GDAC and two MLPs added (dark blue: CANYON-MED, light blue: CANYON-B).

From my perspective, both MLPs do an excellent job in reproducing the measured profile (even CANYON-B, for which no great things should be expected in the Med Sea). Both MLPs even include proper reproduction of variability caused by interleaved water masses along the vertical profile.

However, I see a couple of discrepancies to the manuscript Fig. B1:

1. The measured data is again of different shape, value, and smoothness – for all profiles shown.

The explanation for this behavior is the same as reported in (R2). The smoothness of the measured profiles in Figure B1 differs slightly from the plots reported by the reviewer due to the discretization performed. Regarding the shape and value, the profiles reported in the paper match those in the reviewer's figure, except for the differences caused by the discretization.

2. None of the fine scale features are visible in the manuscript (from what I can eye-ball).

   Again, this is due to the discretization performed, as detailed in (R2).

3. The fine scale features are pretty much mirrored in CANYON-MED; and CANYON-MED seems to match between my figure and manuscript Fig. B1, unlike the measured data. The CANYON-MED MLP is also spot-on on most of the measured profile (in my figure; not in the manuscript Fig. B1).

   The measured profiles reported in the appendix and the figure provided by the reviewer match in shape and bottom values. The only difference is in the discretization, which results in smoother profiles in the appendix.

   Regarding CANYON-MED, the difference arises because we performed the comparison using the first version of the software, which was available when we submitted this paper to the GMD journal. After considering the reviewer's comments, we noticed that the difference in the results is due to the different version of CANYON-MED employed (v1 in our case). Version 2 implemented in Python has recently been released, which provides more accurate predictions due to improvements in its implementation.

   This highlights another strength of the convolutional architecture compared to the MLP architecture: the tuning of hyperparameters in MLPs is more time-consuming and significantly affects overall performance, while CNN can be (at least for some datasets) more resistant to some changes in the architecture and hyperparameters.

4. The float 6903153 in the ION (lower left panel) does not carry a nitrate sensor at all.

   Thank you for pointing out that float 6903153 does not carry a nitrate sensor. After double-checking, we noticed a typo in the float name, we apologize for the mistake. We have updated Figure B1 by changing the float name to the correct one. The correct float cycle is 6901772_145.

These aspects are quite discomforting and raise an eyebrow on whether there are similar discrepancies in other parts of the data, and on which data were compared with what measurements with respect to performance (RMSE) of the various ANNs, both PPCon and others.

After correcting the two errors in reporting the float cycle numbers, we double-checked all other results in the paper. We confirm that the data are correct and accurately reported.

There are a few instances throughout the manuscript, where WMO numbers got a bit scrambled, so it might be as simple as incorrect WMOs and cycles labelled. Or, there may be a more profound issue with the data used, or a mess-up in measured profiles compared to actually different MLP/PPCon predicted profiles. This would be dramatic.

As aforementioned, the problem was solely related to incorrectly reporting the WMO numbers in two cases. After correcting these errors, we have verified that all other data in the manuscript is accurate and correctly reported.

In any case, these issues need to be addressed before a publication.

(a) Where do the discrepancies in the measured data come from? How large is the extent of the discrepancies? What's the impact on the method and its evaluation?

As mentioned previously (R2), the smoother profiles in Figure 3 are due to the discretization performed on these vertical profiles. The vertical profiles of three variables were discretized into 200 points, resulting in a 1-meter resolution for chlorophyll and bbp700 and a 5-meter resolution for nitrate. We acknowledge that this resolution of nitrate, especially in the upper part, led to a smoother appearance compared to the original vertical profile, where the sampling was higher than once every 5 meters.

The other discrepancy that the reviewer noticed where simply due to two WMO which were not reported correctly (middle profiles of Figure 3 and ION profiles of Figure B1). We update the correct WMO on the top of these profiles, and double-check all the others WMO.

Therefore, the extent of the discrepancies is limited to the smoothing effect introduced by averaging the vertical profiles to a consistent resolution. The impact on the method and its evaluation is minimal in terms of the overall performance metrics (RMSE) because the discretization is uniformly applied across all profiles, ensuring that the comparative evaluations remain consistent and fair.

(b) Both from the line of argument presented in this comment, as well as from the illustrations (reproduced Figs. 3 and B1), I cannot recognize that either of claim *a* or *b* can stand. (Rather, MLPs can do a surprisingly good job in reproducing profile shape thanks to (point-wise) water mass characteristics.)
I would therefore ask you to remove those claims entirely, or to substantiate them. (Again, my perspective neglects the Pietropolli et al. 2023a MLP, where I don't have access to the code. It may be that some of your claims *a* or *b* may apply to and be true for the Pietropolli et al. 2023a MLP. But then it cannot take hostage of all MLPs, given that it doesn't apply to other MLP models or architectures.)

We have already addressed the comments regarding these two claims in the earlier part of the review when the reviewer first mentioned these sentences (R1).

(c) Could it be that fine scale characteristics of the vertical profile, like interleaved water masses, cannot be well reproduced by the CNN-based PPCon model/architecture? Because it puts more emphasis on the entire/large-scale profile shape and 'neglects' the information from the water mass properties, which is the only information available to MLPs?

We acknowledge this observation and agree that CNNs and MLPs have different strengths.

Our CNN-based approach emphasizes the overall shape of the vertical profile, which is advantageous for predicting variables like chlorophyll and bbp700. These variables benefit from the CNN's ability to learn and reproduce smooth, consistent profiles. For nitrate, which has more regular shape profiles, both CNNs and MLPs show similar performance.

It is important to note that the smoother appearance of our CNN profiles is primarily due to the discretization applied. We used consistent discretization to ensure uniform input and output dimensions for the convolutional architecture. If we were to increase the resolution of discretization, our CNN would likely be able to reproduce fine-scale characteristics more accurately.

We see the CNN approach as complementary to MLPs, rather than superior. Each method has its advantages depending on the specific variable being predicted. For fine-scale details, increasing the discretization for CNNs could enhance their ability to capture these features effectively, similar to MLPs.

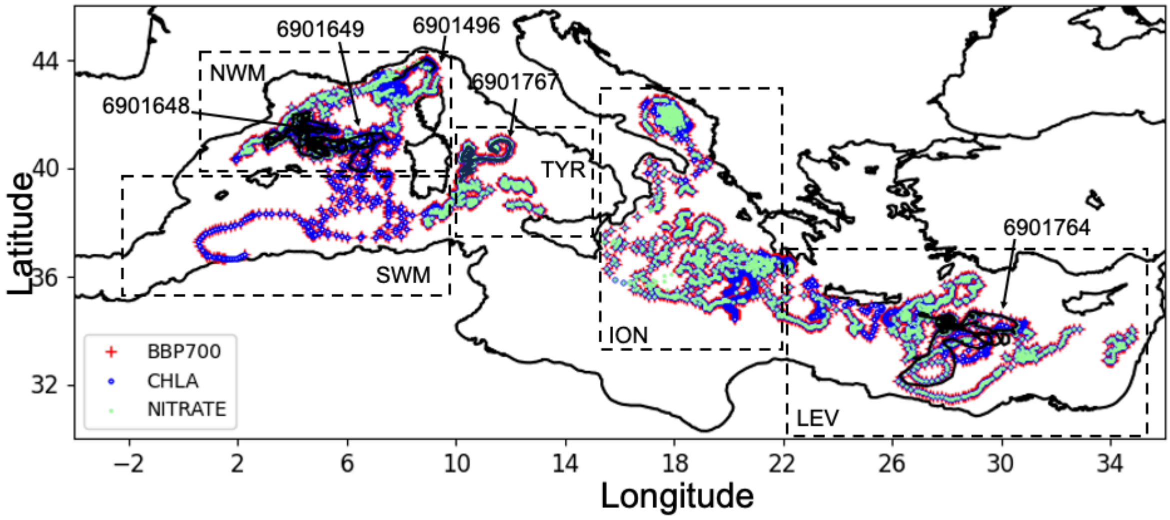
(d) Why are the measured data in the manuscript without fine scale and with mostly eroded mixed layer base?

Since CNNs require input vectors of the same length, we used a 5-meter average in the pre-processing step, as explained in (R2). This averaging smooths the profiles, resulting in the loss of fine-scale features and an appearance of an eroded mixed layer base in the measured data. This pre-processing was applied to input data of both MLP and CNN models, keeping the comparative analysis consistent and fair.

Further points:

- Figure 1: I cannot find float 6901767, the second nitrate float used later on (Fig. 7 and Tab. 8). Or is it a WMO mix-up there (and it's 6901648??).

We corrected and updated the Figure 1.



- l. 133: “Table 2” referred to is missing.

Thank you for pointing out the missing reference to “Table 2” on line 133. The table in question is located on the previous page of the manuscript and reports the components of the PPCon architecture. We will ensure that the reference is clearly indicated.

- l. 134: From the description of Amadio et al. 2023 (and the above experience) it reads that the float data were treated with the “bit.sea python package (Bolzon et al., 2023)”, including “a smoothing task”. For a manuscript that focuses and puts emphasis on profile shapes, this is important information that belongs into the description of the dataset (and cannot be hidden somewhere inside some reference). And which warrants some (brief) discussion, because it seems that your figure’s mixed layer bases are much more eroded than in the original float data. Which may have implications if PPCon were to be used for augmenting a float dataset.

In the case of the present work, we already explained that we performed a 5-m discretization to have a 200-point length profile for nitrate.
In the case of Amadio paper, the smoothing is done to interpolate the float profiles (either BGC-Argo or reconstructed with MLP) to the layer thickness of the model vertical discretization which varies with depth (about 1.5m at the surface, 5m at a depth of 40m, 10m at a depth of 170m, 25m at a depth of 1000m)
The two tasks respond to different needs.

- section 3.1: The question remains on why such a complex architecture was chosen and what should be gained from 4 separate MLPs per singular input (over replication of input data into a 200x1 vector each) to be fed into the CNN. But I won’t insist and just note that this is still weakly motivated.

We thank the reviewer for the comment. We will keep it in consideration for further developments of PPCon.

- l. 182: “adding zero padding to the borders of the input tensor”: This requires the input to be normalized, i.e., mean data subtracted (and ideally divided by the standard deviation). A batch normalization is mentioned only for the output tensors in the text, not for the input tensor. From table 2 I understand that batch normalization (BN) is done for every layer’s input, is it? So maybe follow the
logic as it is outlined in the table also in this paragraph: Layers consist of BN, SELU and have Dropout.
To compress/decompress information we have kernels/strides, which require padding with …

Thank you for your feedback regarding the normalization and padding of input tensors. You are correct in noting that normalization is essential for ensuring that zero padding is effective.

As outlined in Table 2, batch normalization (BN) is applied to the input of every layer. This includes the input tensor, ensuring that it is normalized by subtracting the mean and ideally dividing by the standard deviation.

- l. 314: “For the nitrate variable, the reconstruction performed by the MLP model is also reported (Pietropolli et al., 2023a).” No, it’s not – but it should! :-)
For Figure 3 and also Figure B1, I’d also suggest to use a bit wider spectrum of colour than different shades of blue, to be able to better distinguish the different models.

Regarding the statement on line 314, we intended to convey that the reconstruction performed by the MLP model for the nitrate variable is reported in the appendix. The sentence will be changed as follows:

"For the nitrate variable, the reconstruction performed by MLP models (Pietropolli et al., 2023a and Fourier et al., 2XXX) is also reported in appendix B"

- l. 317f: “...than the previous MLP architecture by Pietropolli et al. (2023a), but similar to MLPs by Fourrier et al. or Bittig et al.”? But my general recommendation would be to drop the second part of the sentence.

Regarding the sentence on line 317, we agree with your recommendation and will drop the second part of the sentence to improve clarity and focus.

- l. 472: That predictions tend to better in deep waters compared to surface waters is true for of all approaches; it’s not unique to PPCon. It’s a feature due to the different variability of the Ocean’s parameter space at depth vs. at the surface, nothing of a particular model’s architecture.

We agree with the reviewer's comment. Improved prediction accuracy in deep waters compared to surface waters is a common characteristic of neural network approaches due to the inherent variability in the Ocean's parameter space.

However, we also aim to highlight that our approach, PPCon, demonstrates better performance in deep waters compared to other models. The difference in performance between PPCon and MLP models is more pronounced in deep waters, reflecting the enhanced capability of PPCon in these conditions.

- Fig. B1 caption: 6903153 in ION has no nitrate sensor.

As said in previous response, after double-checking, we noticed that the profile reported does not correspond to the profile contained in the image. We apologize for the mistake in reporting the float profile number. We have updated Figure B1 by changing the float name to the correct one. The correct float cycle is 6901772_145.

Minor points:
- l. 42: replace "frequency" by something more fitting? Maybe
- l. 133: "Table 2" referred to is missing.
- l. 145: sample date?
- l. 303: "in the ION[+, SWM, and TYR,] with RMSE values below 0.5 [+unit]"?
- l. 306f.: check "which are the highest ..." – Does this still apply?
- l. 395: "DT" was this abbreviation introduced? Why not spell it out

Typo's:
- quite a few instances: "BCG" instead of "BGC"
- l. 41: closing ")"
- l. 263: α*s* ?
- l. 377: closing ")"
- l. 403: understand
- l. 439; BGC-Argo network
- Fig. B1 caption: 6902904
- l. 102/l. 509: Cross-check the Bittig et al. 2019 reference. That's the one you intended? (technical documentation vs. a published manuscript?)

We appreciate the reviewer's feedback on these points and typographical errors. We have addressed all these issues in the revised manuscript.