Reviewer Comment #2

Report for the manuscript

"A global monthly field of seawater pH over 3 decades: a machine learning approach" by G. Zhong et al.

The manuscript presents an application of machine learning techniques to reconstruct global fields of seawater pH covering the years 1992 to 2020 at 1° and monthly resolutions. This research complements to the exiting research studies on generating pH values at the surface ocean by providing global maps of seawater pH for depth levels up to 2000 m, essential for understanding ocean acidification and its impacts on ecosystems in the ocean interior.

General comments:

Machine learning offers flexible frameworks to link in situ observations of marine carbonate system variables with relevant environmental conditions. The novelty of this research lies on the idea of mapping on direct pH data from the surface down to different depth layers instead of computing through the carbonate system speciation given prior parameters as pCO2, DIC, …. The authors are also able to integrate various data sources, including in-situ measurements and satellite-based datasets as input of pH estimates.

While the manuscript presents the effectiveness of machine learning models (SOM, Stepwise FFNN, FFNN), there is a crucial need for greater transparency regarding model architecture, selection process of predictors, hyperparameter tuning, model accuracy and uncertainty quantification. Including these details would enhance the reliability of the proposed product and interpretability of the study.

Response: Thanks for the suggestion. We have revised the methodology section for better clarity, the detailed changes are as the following responses.

1.Methodologies:

Even interpolation with direct pH data may be feasible but model accuracy remains limited due to the modest amount and data coverage of GLODAP pH used for training and validation, that may be much more problematic at the deep sea. More statistics on GLODAP data used for FFNN training and validation are needed for a comprehensive evaluation on the model efficiency. It would be worthy to test the proposed method by subsampling with the same GLODAP tracks on the exiting products that offer global maps of pH at some levels depths (see for instance: https://doi.org/10.48670/moi-00015).

Response: Thanks for the suggestion. As we only divided two vertical layer to train FFNNs, the number of training samples at the deep sea is comparable with the mixed layer near surface, and only the depth coverage of these training samples was notably larger than the mixed layer. Although these training samples were sparser in space, the FFNNs suggested a low predicting pH at the deep sea, which may be caused by the smaller seasonal and annual pH variability at the deep sea.

We used about 75% of GLODAP data for training and about 25% for testing in each iteration of evaluation, with training and testing groups divided by years. After repeating four times and changing testing groups in each iteration, all GLODAP sample has been used for testing once to carry out a comprehensive evaluation. We have added a figure showing the statistics of training and testing sample in the supplement as the following Figure S1.

The mentioned product in https://doi.org/10.48670/moi-00015 starts from 2021, which does not coincide with the temporal coverage of our product and were not used for evaluation. Furthermore, other existing machine learning-based products are only available for the global surface ocean, so we only presented comparisons of surface pH trends with these products. For further evaluation of our product in deep sea, we have added comparison with qualified pH data from Biogeochemical Argo float dataset in the validation section.

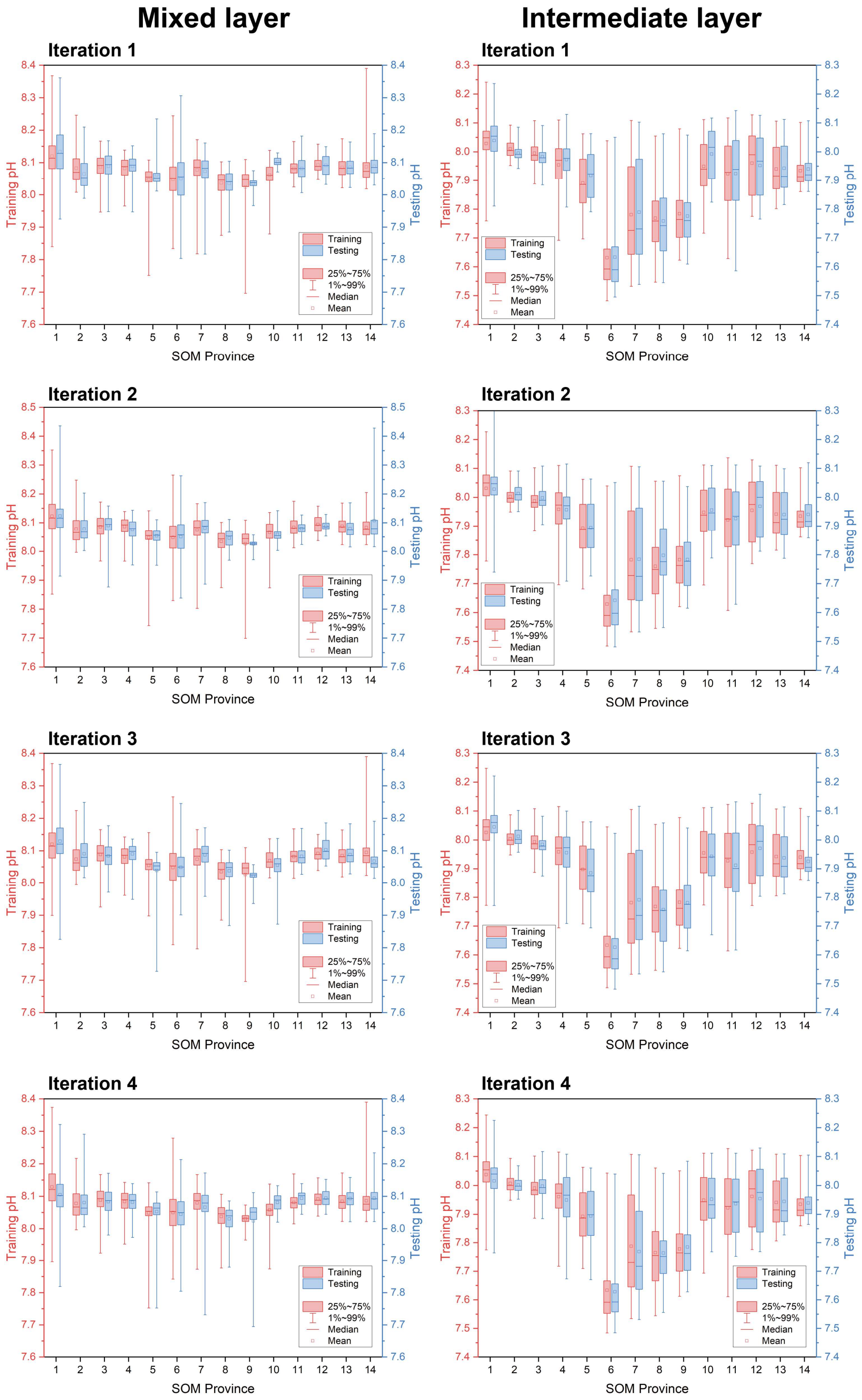

Figure S1. Statistical distribution of GLODAP samples used for training and testing in each province. Iteration 1-4: repeated evaluation with different training and testing samples dividing by years. Samples in 1992, 1996, …, 2020 were used for testing and the rest were used for training in iteration 1; samples in 1993, 1997, …, 2017 were used for testing and the rest were used for training in iteration 2.

The use of all the three ML algorithms seems redundant and would gain uncertainty for pH mappings (particularly with limited number of training data). FFNN itself would be skillful enough to interpolate pH without SOM for regional clustering and stepwise FFNN for predictor selection (see for instance: Broullón et al., 2019; 2020; Chau et al., 2022; 2024). Results derived from SOM shown in this study (e.g. Figures 1 and 8) does do not reflect the sharing patterns of mechanisms in CO2 uptake and storage, e.g., at temperate (equatorial) zones between different oceanic basins. pH estimation based on clustered biomes creates discontinuity at the regional boundary. Clustering and predictor selection are probably available in FFNN training phase as it automatically weights and select input neurons to get optimal output. In addition, the use of lat, lon, depth as predictors potentially locallizes training data and would allow to interpolate pH at non-observed locations. Results after Stepwise FFNN are not accurate enough (see comment for Table 2 in Specific Comments).

Response: Thanks for the suggestion. SOM-based regional clustering has also been proved effective in reducing regional predicting error in machine learning mapping of carbonate system variables, such as Landschützer et al. (2016), Iida et al. (2021), and Zhong et al. (2022). Although there is discontinuity problem, we decided to use SOM clustering based on the following main reasons:

(1) It is difficult to adjust the FFNN architecture and input predictor to obtain the optimal performance in different regions and depths. The number of neurons and combination of predictors adjusted to reduce predicting error in specific region may also lead to higher errors in other regions. Using regional-specific predictors and model architectures can reduce errors in different regions simultaneously.

(2) Inputting over 300 thousand unbalanced GLODAP samples into one FFNN model may generate biased outputs in regions with sparser samples. Previous research suggested the one FFNN trained with unbalanced samples will generally output values biased toward the majority pattern of training samples (Zhong et al., 2024). The one FFNN model only get optimal output in most data-rich areas, such as north Pacific and north Atlantic with the data amount far more than other areas. The output pH value in data-sparse areas may be more biased toward pH pattern in data-rich areas.

(3) The discontinuity problem can be solved with the further accumulate of pH measurements and

improvement in SOM technical. In the surface ocean, the discontinuity problem did not appear in most boundaries.

(4) Training only one FFNN is also feasible, but is preferable in the method that mapping DIC and TA first and then calculating other variables, as these two variables are relatively more conservative.

Sample position and time were used as predictors because the collected environmental variables are not enough comprehensive to cover all ocean processes affecting pH, as gridded products of many variables are currently not available or only climatological. If more environmental variables are included in future works, the sample position and time will fail to compete with other environmental variables in the predictor selection procedure.

Iida, Y., Takatani, Y., Kojima, A., & Ishii, M. (2021). Global trends of ocean CO_2 sink and ocean acidification: an observation-based reconstruction of surface ocean inorganic carbon variables. *Journal of Oceanography*, *77*, 323-358.

Landschützer, P., Gruber, N., & Bakker, D. C. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, *30*(10), 1396-1417.

Zhong, G., Li, X., Song, J., Qu, B., Wang, F., Wang, Y., ... & Duan, L. (2022). Reconstruction of global surface ocean p CO 2 using region-specific predictors based on a stepwise FFNN regression algorithm. *Biogeosciences*, *19*(3), 845-859.

Zhong, G., Li, X., Song, J., Wang, F., Qu, B., Wang, Y., ... & Dai, J. (2024). The Southern Ocean carbon sink has been overestimated in the past three decades. *Communications Earth & Environment*, *5*(1), 398.

Please consider to provide more details on the data preprocessing steps, including how different data sources were transformed and harmonized, as well as elaborate on (1) the choice of many predictors (e.g., why temperature anomaly was used instead of the standard ones; why both number of months and Year Month are needed,...), (2) machine learning algorithms used, and (3) the rationale behind selecting them over other potential models.

Response: The collected products from different sources are all gridded datasets and the most are in the same 1° resolution. In the preprocessing step, products with higher resolution were converted into 1° resolution by averaging all data within the same 1° grid into one value.

The predictors were selected from collected environmental variables and sample information, which are expect to be as many as possible and cover most ocean processes affecting pH. Most variables have been used for reconstruction of ocean carbonate system variables in previous researches, such as temperature and its anomaly. Some variable seems redundant but have different features. For example, the sample time has no seasonal cycle pattern information when using the number of months and Year, and is disconnected between years when using Months as 1-12. Therefore, we collected as many variables as possible and then selected predictors using a Stepwise regression algorithm based on FFNNs (referred as Stepwise FFNN), according to the pH predicting errors when using different combination of variables as FFNN inputs. This algorithm has been proved to have capacity to identity the most informative variables in previous pCO$_2$ mapping research and can effectively reduce regional predicting errors. We have revised the section 2.3 pH product construction for better clarity of the predictor selection and product construction procedure.

2.Model robustness and Uncertainty Analysis:

Although the authors have shown the validation against independent datasets from GLODAP, other ML-based methods, and some time series of direct measurements (e.g., HOT, ESTOC, BATS), the examination limited almost for the surface ocean. Evaluation and analysis of model accuracy at seawater depths would provide a clearer understanding of the model performance. Some timeseries stations have offered pH data below the surface (e.g., HOT, ESTOC, BATS), other sources for data evaluation can be found in Sutton et al., (2019), Lange et al., (2024).

Response: Thanks for the suggestion. We have added the BAT, HOT, and DYFAMED time series data below the surface in the validation section.

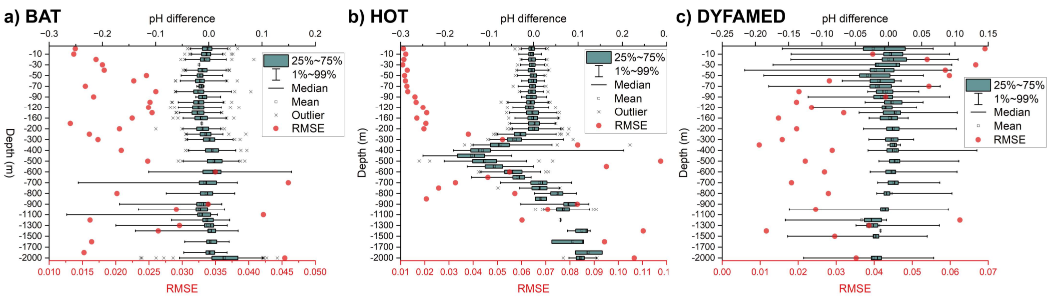

Figure 7. RMSE and pH difference between FFNN pH and time series observations at different depths. a) BAT station at 31°50' N, 64°10' W based on data from 1992 to 2020; b) HOT station at 22° 45' N, 158° 00' W based on data from 1992 to 2020; c) DYFMED station at 42.3°N, 7.5° E based on data from 1998 to 2017.

Despite higher RMSE at certain depths, the RMSE at most depths in the deep areas of the BAT station and DYFAMED station is below 0.03, indicating that the notable deviations may only occur at local scale. This notable deviation may be due to sparser GLODAP measurements in certain areas, or difference in depths between pH product and independent observations. For example, the FFNN pH product was reconstructed at depths of 1800 m and 2000 m in the bottom. If the time series observation is at 1910 m depth, it will be compared with the FFNN pH value at 2000 m in the independent validation. This depth difference significantly increases the pH error in validation based on independent data if the number of independent observations was limited.

The manuscript would benefit from a more comprehensive uncertainty analysis. It is not convinced that adding the source of RMSE from FFNN [H+] is essential for the quantification of pH uncertainty as the author ultimately produce 4D pH fields from GLODAP pH data (not [H+]. Presenting a comparison between RMSE of pH derived from direct GLODAP pH and from [H+] in the supplementary is more appropriate. Instead, it would be worthy to consider predictors' uncertainty in the total uncertainty of reconstructed pH.

Response: Thanks for the suggestion. Adding the source of RMSE from FFNN $[H^+]$ is mainly to distinguish the uncertainty between areas with same pH RMSE but different pH levels, as the uncertainty would be the same between these areas if calculated directly based on pH RMSE. The section of comparison between pH and [H+] has been moved to the supplementary. Including all predictors' uncertainty will be a better way to estimate the pH product uncertainty. However, as described in the supplementary section *Uncertainty and construction method of selected ocean products*, the uncertainty of particular predictor products is unclear. It is not feasible to convert the uncertainty of predictor products through the FFNN when some inputs are missing. Therefore, we have to estimate the pH product uncertainty in a different way.

Specific Comments:

Lines 44-45: The two studies report fast decrease in pH in the ocean interior. The authors should reword this sentences.

Response: The fast decrease in pH was reported in subsurface above 500 m in these two studies, but a relatively slow acidification can be also observed in deeper areas. The sentence has been corrected as the following:

"Meanwhile, relatively slow acidification was found in the deep Atlantic Ocean below 2000 m (Guallart et al., 2015), and rising pH in deep waters around 1000 m was also reported in the North Pacific Ocean (Ishizu et al., 2021)."

Lines 45-46: Can be rephrased (for instance: "there remains a need to enhance our understanding of global ocean acidification rates across varying depths.")

Response: Thanks for the suggestion. The sentence has been modified as the following:

"With limited reports about acidification below the surface, there remains a need to enhance our understanding of global ocean acidification rates across varying depths."

Line 49: “the global mapping of” to “global reconstructions of”.

Response: The sentence has been corrected following the suggestion.

Lines 50-51: Gregor and Gruber, (2021) and Chau et al., (2024) have published full datasets of many carbonate variables (pH, DIC, Alkaninity included).

Response: The sentence has been modified as the following:

"Recent applications of machine learning methods in global reconstructions of marine carbonate system variables have facilitated global-scale research on the acidification and carbon cycle, including the single/ensemble-based FFNN method and the SOM-FFNN method for mapping surface ocean partial pressure of CO_2 (pCO_2, Landschützer et al., 2014; Chau et al., 2022; Zhong et al., 2022; Chau et al., 2024), dissolved inorganic carbon (DIC, Broullón et al., 2020; Keppler et al., 2020; Gregor and Gruber, 2021; Chau et al., 2024), and alkalinity (Broullón et al., 2019; Gregor and Gruber, 2021; Chau et al., 2024)."

Lines 55-56: “The construct pH product”; this term is not a standard scientific or technical term. Maybe replace with "the proposed product" or "the recontruced pH data".

Response: The term "The construct pH product" has been replaced by "the proposed pH product".

Line 61: (Lauvset et al., 2022) → update refs for GLODAPv2.2023 and cite right after 2023 version instead.

Response: The citation has been updated.

Line 62: “in-situ temperature”; Reviewer do not see any specific role of temperature in pH mappings throughout the manuscript.

Response: The in-situ temperature was mentioned here is only for indicating that this product is not at 25°C, as the GLODAP dataset also provides pH data corrected to 25°C.

Line 63: Table 1 show most of predictors’ products with no depth levels, it is not clear how to map pH with constant values of predictors over depths!

Response: The product of temperature, salinity, nutrient concentration, dissolved oxygen, DIC, and alkalinity provide values across different depths.

Lines 66-70: many predictors have been used but there are no hints (citations) showing why they should be included in model fitting.

Response: The citations of previous application of the predictors have been added.

Line 80: “such as the ocean currents product” → should be at the beginning of the sentence.

Response: This term has been moved to the beginning of the sentence.

Table 1: please mention temporal and vertical resolutions of predictors’ products used; comments on how to derive transformation for Date (Year, Month,...).

Response: The temporal and vertical resolution have been added in Table 1. Product with daily or weekly resolutions were converted to the monthly resolutions by directly averaging all values within the same month, and there is no product with year resolution.

Line 116: “Therefor”.

Response: The typo has been corrected.

Figure 1: consider to not use SOM (see in General comments).

Response: Using one FFNN model can indeed reconstruct global gridded pH data, but this method does not account for the regional differences in factors affecting pH. The primary feature of our method is the consideration of regional differences in pH drivers, allowing for the selection of the most suitable predictors for pH reconstruction in each region, thereby increasing accuracy. Despite continuity issues at some boundaries, the overall reconstruction error is lower than using one model. In our previous research on reconstructing gridded $p$$CO_2$ data, we also employed the same SOM method to use regional-specific predictors, which significantly reduced reconstruction errors and was well-received by peers.

Lines 153-154: “To mitigate the influence of the FFNN's initial state on predicting values, multiple networks with the same structure but different initial states were trained and their results were averaged”; standard deviation from output averaging should be reported.

Response: The figure showing mean standard deviation between FFNN pH with different initial status has been added in the supplement as the following:

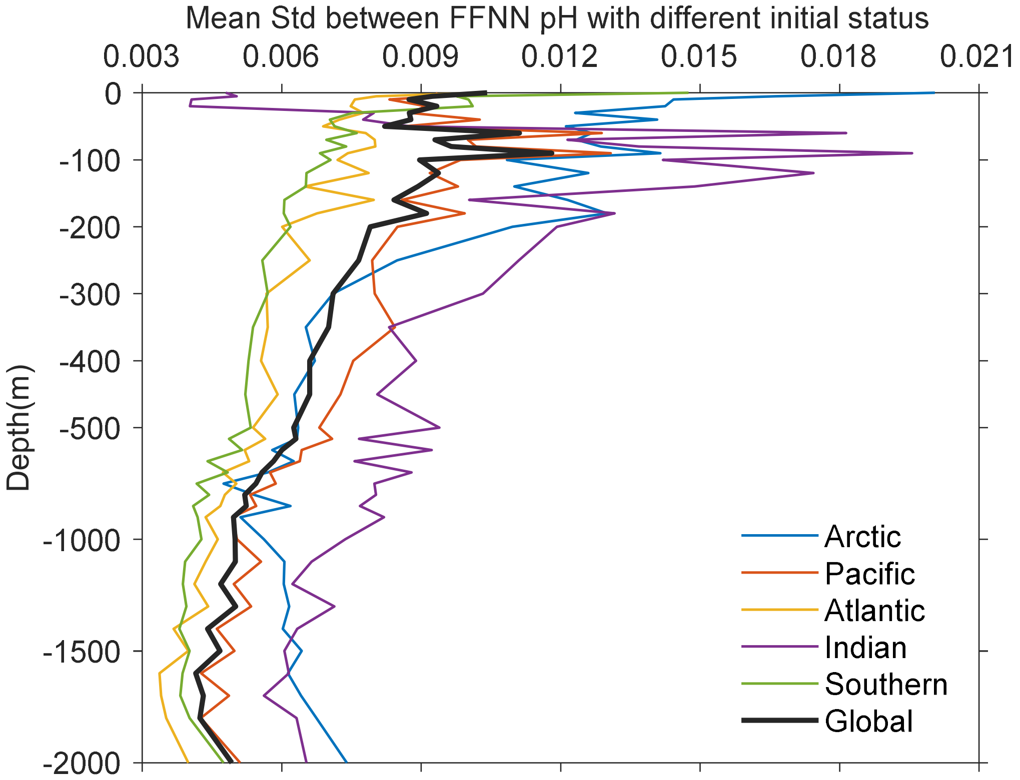

Figure S5. Mean standard deviation between FFNN pH with different initial status.

Tables 2, 3: consider to not use Stepwise FFNN to select predictors. FFNN itself would choose which inputs benefit model training. Results from Stepwise FFNN are not align with marine CO2 system features and driving mechanisms. For instance, temperature is one of the key factors modulating CO2 absorption over the Arctic and subpolar regions (thus performs impact on pH); however, this predictor was not chosen after Stepwise FFNN. In addition, please clarify the use of both temperature (salinity,...) and their anomalies that are redundant information and may challenge FFNN training!!!

Response: In current commonly used methods for reconstruction of global ocean gridded data, the predictors are typically selected empirically. The manually selected predictors vary in different methods for constructing the same marine chemical variable, contributing to partial differences between the products. Our algorithm selects input predictors based on statistical characteristics, eliminating the influence of subjectivity and randomness of manual selection. In some regions, effective predictors for reducing pH reconstruction errors may differ from predictors identified based on experience. For example, in the Arctic Ocean, temperature is not used because its effect on pH is already reflected in other input predictors like $p\mathrm{CO_2}$, which is notably affected by temperature. The pCO_2 increases with the increasing temperature. Additionally, the seasonal temperature variation in the Arctic Ocean is small, so other parameters can sufficiently reflect the effect of temperature. Including temperature as a pH predictor for the Arctic Ocean increased the pH reconstruction error estimated based on GLODAP samples, which is why we excluded temperature when reconstructing pH for the Arctic Ocean.

The monthly anomalies of temperature only reflect seasonal and interannual variations, removing the regional distribution features of temperature. In the 50-60S region of the Southern Ocean, directly using temperature as input predictors primarily reflects the spatial distribution characteristics of temperature, as the scale of regional differences in temperature are much greater than seasonal and interannual variations. This disturbed the model's learning of seasonal and interannual temperature changes. Using both temperature and its monthly anomalies to reflect regional distribution and temporal changes can reduce pH reconstruction errors. Therefore, both temperature and its monthly anomalies are used as pH predictors in the subpolar Southern Ocean. Similarly, other variables are generally not used together with their monthly anomalies in most regions. They are only used together when it can reduce pH reconstruction errors.

Line 174: RMSE; the metric for validation is not consistent with training (MAE, Line 135).

Response: This is because the extreme values have higher weighting if the reconstruction error was represented by RMSE compared to the MAE. In the selection procedure, the algorithm was design to focus on reducing errors for the majority of testing samples rather than particular samples with extreme values. The MAE can also be used to present FFNN performance in the validation section, but RMSE is more commonly used. So we used RMSE in the validation section.

Line 183: Please clarify how to define H+ here.

Response: $[H^+]$ is the molar hydrogen ion concentration here.

Lines 191-206: It's worthy to rework on this section: It is not ease to interpret the uncertainty quantified with pH0 and sigma (see also in General Comments for details).

Response: Although this method is not commonly used, it can better represent the uncertainty of the reconstructed pH products. This method includes the main factors influencing uncertainty, including the pH reconstruction errors and the notable differences in $[H^+]$ that caused by the same pH error at different pH levels. For example, the same pH error of 0.02 lead to a difference in $1.42*10^{-9}$ M of [H^+] when pH is 7.5, and lead to a difference in $1.42*10^{-10}$ M of [H^+] when pH is 8.5. The difference in [H^+] differs notably but the uncertainty directly estimated by pH errors will be the same.

Lines 212-214: “A better performance of the FFNN was found in the intermediate layer, with testing samples more concentrated on the y=x line. The RMSE in the mixed layer is 0.034, higher than 0.026 in the intermediate layer.” Reviewer would expect to see reverse results: there exist very few pH data and predictors’ information which are able to support pH estimation in the deep sea than the shallower layer !!! Reconstruction errors (Uncertainty) would much higher there than the surface.

Response: Although the pH measurements are much sparser in the deep sea, we trained the FFNN using all samples in the deep sea from mixed layer depth to 2000 m. The number of training samples in the deep sea is even more than that in the mixed layer, as covered depths are much broader. Additionally, the FFNN underestimation of seasonal amplitude and short-term fluctuations is the main source of reconstruction errors, which are notably smaller in the deep sea. This can be observed in the validation based on time series station, the reconstruction error is notably higher when pH seasonally peaked and troughed. The same pattern of decreasing RMSE with depth can also be observed in the 3D reconstruction of DIC and TA in previous research, such as Broullón et al., 2019 and Broullón et al., 2020.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., ... & van Heuven, S. M. A global monthly climatology of total alkalinity: a neural network approach. Earth System Science Data, 11, 1109-1127, https://doi.org/10.5194/essd-11-1109-2019, 2019.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., ... & Kozyr, A. A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach. Earth System Science Data, 12, 1725-1743, https://doi.org/10.5194/essd-12-1725-2020, 2020.

Line 219: remove "predicting" here and elsewhere, FFNN pH is informative enough.

Response: The term "predicting" has been removed in all contexts.

Lines 224-225: Any clarifications to have errors lower in the deep sea than the surface. I would appreciate of any clarification.

Response: In the upper ocean above 2000 m, the underestimation of seasonal amplitude and short-term fluctuations has greater impacts on reconstruction errors than that of sparse training samples in the deep sea. Therefore, the validation based on the GLODAP dataset show a decreasing RMSE with increasing depths.

Line 230: “predicting error” is not correct. Please use "prediction error" or "reconstruction error" instead.

Response: Thanks for the suggestion. This term has been replaced by "reconstruction error".

Lines 237-238: “The RMSE in the early years was relatively higher than in recent years, while the number of GLODAP measurements increased with the years (Figure 5c)”. Adding curves for number of GLODAP pH in each subplot will help to evidence the statements for Fig 5.

Response: The number of GLODAP pH has been added in Figure 5c.

Lines 249-: Why do the results show for Stepwise FFNN while final reconstruction is done with FFNN?

Response: The reconstruction in this work was based on a two-step method that including the predictor selection by stepwise regression using FFNN and the FFNN fitting of non-linear relationship. So, we named the pH product as the Stepwise FFNN product to summarize the whole method.

Lines 251 - 254:

“The surface seawater pH of our Stepwise FFNN product decreased by 0.0017 ± 0.0007 yr-1 on average during the past three decades at the BAT station, close to the -0.0018 ± 0.0001 yr-1 of BAT time series observations in the same period (Bates et al., 2020). At the ESTOC station, the Stepwise FFNNN product and time series observations were also well consistent, with the RMSE of only 0.009 and a similar long-term trend (Chau et al., 2022).” these quotations are not correctly mentioned in refs (Bates et al., 2020; Chau et al., 2022). The pH decreasing rate was about 0.0019 ± 0.0001 per yr over the period 1983-2020 in the former study. Please clarify that the authors have used the data to compute the trends by themselves. Furthermore, Chau et al., (2022) do not provide long-term trend estimates of pH but Chau et al., (2024).

Response: The citation has been corrected to González-Dávila et al. (2010) for ESTOC and Chau et al., (2024) for the CMEMS pH product. The pH trends from previous products in Table 4 were computed using data from 1992 to 2020 in the BAT and HOT station, and were computed using data from 1995 to 2010 in the ESTOC station, to eliminate the influence of different temporal coverage. The description of compute period has been added in remarks of Table 4.

Line 255: remove "only" in this paragraph and elsewhere as 0.01 in pH is indeed large correspondingly to a difference in 26% of H^+ (acidity level)".

Response: Thanks for the suggestion. The word "only" has been removed.

Table 4: report uncertainty estimates for all other products.

Response: The uncertainty of other products has been added.

References

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., González-Dávila, M., Jeansson, E., Kozyr, A., and van Heuven, S. M. A. C.: A global monthly climatology of total alkalinity: a neural network approach, Earth Syst. Sci. Data, 11, 1109–1127, https://doi.org/10.5194/essd-111109-2019, 2019.

Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., Santana-Casiano, J. M., and Kozyr, A.: A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach, Earth Syst. Sci. Data, 12, 1725–1743, https://doi.org/10.5194/essd-12-1725-2020, 2020.

Chau, T. T. T., Gehlen, M., and Chevallier, F.: A seamless ensemble-based reconstruction of surface ocean pCO2 and airsea CO2 fluxes over the global coastal and open oceans, Biogeosciences, 19, 1087–1109, https://doi.org/10.5194/bg-19-10872022, 2022b.

Chau T T T, Gehlen M, Metzl N, Chevallier F (2024). CMEMS-LSCE: A global 0.25-degree, monthly reconstruction of the surface ocean carbonate system, Earth System Data, 16, 121–160, https://doi.org/10.5194/essd-16-121-2024.

Lange, N., Fiedler, B., Álvarez, M., Benoit-Cattin, A., Benway, H., Buttigieg, P. L., Coppola, L., Currie, K., Flecha, S., Gerlach, D. S., Honda, M., Huertas, I. E., Lauvset, S. K., Muller-Karger, F., Körtzinger, A., O'Brien, K. M., Ólafsdóttir, S. R., Pacheco, F. C., Rueda-Roa, D., Skjelvan, I., Wakita, M., White, A., and Tanhua, T.: Synthesis Product for Ocean Time Series (SPOTS) – a ship-based biogeochemical pilot, Earth Syst. Sci. Data, 16, 1901–1931, https://doi.org/10.5194/essd-16-1901-2024, 2024.

Sutton, A. J., Feely, R. A., Maenner-Jones, S., Musielwicz, S., Osborne, J., Dietrich, C., Monacci, N., Cross, J., Bott, R., Kozyr, A., Andersson, A. J., Bates, N. R., Cai, W.-J., Cronin, M. F., De Carlo, E. H., Hales, B., Howden, S. D., Lee, C. M., Manzello, D. P., McPhaden, M. J., Meléndez, M., Mickett, J. B., Newton, J. A., Noakes, S. E., Noh, J. H., Olafsdottir, S. R., Salisbury, J. E., Send, U., Trull, T. W., Vandemark, D. C., and Weller, R. A.: Autonomous seawater $p\mathrm{CO_2}$ and pH time series from 40 surface buoys and the emergence of anthropogenic trends, Earth Syst. Sci. Data, 11, 421–439, https://doi.org/10.5194/essd11-421-2019, 2019.