Reviewer#2

First of all, I would like to thank the authors for the careful and thoughtful responses to my comments and suggestions. I believe their revisions have improved the manuscript, and the new details make it clearer to read. However, there are still some issues that deserve to be addressed before the manuscript can be published.

Response: Thanks for your suggestions, which help improve the quality of our manuscript. We have revised the manuscript according to the raised remaining issues. The detailed changes are as the following responses.

Specific points to raise :

To avoid confusion regarding temperature, I suggest adding the following clarification: *"Here, we present a monthly four-dimensional 1°×1° gridded product of global seawater pH at total scale and in-situ temperature (without standardization to 25°C)."*

This will ensure that readers understand the product focuses on pH measurements, while avoiding any implication that it is also a temperature product. I recommend adding this clarification at line 60 and elsewhere when necessary.

Response: Thanks for the suggestion. We have added "(without standardization to 25°C)" following the description of in-situ temperature in the abstract and method section.

- The use of sin(Lat) as a predictor is questionable since latitude is not circular.

Response: This normalization method was inspired from previous research, such as Denvil-Sommer, A., et al. (2019), where they also normalized latitude and longitude to radians using sine and cosine transformations. Also, we have corrected the description name in Table 1 to "Sine of (latitude · π/180°)", "Sine of (longitude · π/180°)", and "Cosine of (longitude · π/180°)". As we used the "sind" and "cosd" function (sind(latitude) equals sin(latitude · π/180°)) in MATLAB, the original description was misleading and has been corrected.

The use of sin(Lat) as a predictor remains questionable, as latitude is not a circular (or periodic) variable. The response mentions that this normalization method was inspired by previous research, such as Denvil-Sommer et al. (2019), and the correction to the description in Table 1 has been made. However, I still have concerns for the following reasons:

• Sine and cosine functions are typically applied to periodic variables, such as longitude or day of the year, where values "wrap around." Latitude, on the other hand, is not periodic—*lat = -90°* is not equivalent to *lat = 90°*.

• Furthermore, the expression *sin(lat · π/180°)* seems inappropriate, as radians conversion should account for the full range of latitude values. If anything, it would be *sin(lat · π/90°)*, but even this is not ideal for latitude.

Given these issues, I maintain that latitude is not a suitable candidate for inclusion via sine and cosine transformations.

Response: Using the sine function is not for connecting the poles but rather normalizes the latitude to the

range of [-1, 1] to be consistent with the scale of transferred longitude. Therefore, we use $\sin(\text{lat} \cdot \pi/180°)$ instead of $\sin(\text{lat} \cdot \pi/90°)$, so that after transformation, lat = -90° becomes -1 and lat = 90° becomes +1, which are not the same. The reason for using the sine function is that the relationship between latitude and many natural processes, such as solar radiation and temperature, is typically nonlinear. The NOAA Greenhouse Gas Marine Boundary Layer Reference xCO_2 product used as a predictor was also based on sin(Lat) (Lan et al., 2023). Using the sine function better captures the nonlinear variation of pH and its influencing factors with latitude. Furthermore, using sin(lat) also incorporates additional information about the variation of grid area with latitude. The area of the 1° grid in high-latitude regions is significantly smaller than that in low-latitude regions, which is not indicated by latitude. When using sin(Lat), the difference in sin(Lat) between grid points is linearly related to the grid cell area, providing a better representation of the spatial information of used GLODAP pH observation data. Therefore, we used the sine function to standardize the latitude.

Lan, X., Tans, P., Thoning, K., & NOAA Global Monitoring Laboratory. NOAA Greenhouse Gas Marine Boundary Layer Reference - CO_2. [Data set]. NOAA GML, https://doi.org/10.15138/DVNP-F961, 2023.

-   Clarify how depth is used as a predictor and whether it corresponds to the depth of retrieval of the output or if the FFNN estimates X values for X depth levels.
Response: Thanks for the suggestion. Depth was used in the same way as latitude or time-related variables in Table 1. The sample depths of GLODAP measurements were input into FFNNs during the training process, and the depths of 41 depth layers defined as target output layers were input into FFNNs during the interpolation process to generate a product covering 0-2000m. The description has been added in the 2.1 section as the following: "Temporal and spatial sample information, including latitude, longitude, depth and sample time, was also used as supplementary variables. Latitude and longitude were normalized to radians using sine and cosine transformations, to present connected sample position information. The spatial sample position and time information of GLODAP measurements were input in the training of FFNNs, and the spatial position and time of defined 1° and monthly product grids were input into FFNNs during the interpolation process to output a gridded product."

Thank you for the explanation regarding how depth is used as a predictor in the FFNN models. However, I am still unclear as to why pressure (*pres*) is not systematically included as a predictor in every FFNN (as seen in Table 2). Given that depth-related information is a critical factor, especially in oceanographic models, it seems logical that *pres* would be consistently used alongside depth.
Additionally, this raises the broader question of why other key spatial-temporal predictors, such as longitude, latitude, and time, are not always systematically included as inputs in the FFNNs. It's unclear why time, in particular, is only integrated in some models and not others, given its fundamental importance in understanding temporal variability in the data.
I suggest providing a clearer rationale for the selective use of these variables and ensuring that key predictors are consistently applied across all FFNNs, or explaining why their inclusion is sometimes omitted.
Response: Thanks for the suggestion. The reason for not using pressure is the high correlation between

pressure and depth. Different from the fundamentals of physical or biogeochemical ocean models, the FFNN method is based on a non-linear relationship regression between output and input environmental variables. For the FFNN method, more input variables do not always mean better, which is also the reason for performing a predictor selection procedure before the reconstruction in this work. The input predictors with high co-correlation tend to cause FFNN overfitting and worse performance. For example, in previous research reconstructing the vertical profile of seawater alkalinity, the pressure and depth are also not used together (Broullón et al., 2019). Therefore, to avoid overfitting we did not use pressure alongside depth.

In certain regions, latitude, longitude, and time are not included as predictive parameters because they do not provide sufficient information or effectively reduce the FFNN pH predicting error. While latitude, longitude, and time are associated with variations in pH, they do not directly influence the spatial or temporal distribution pattern of seawater pH. Instead, these variables are only related to regular zonal distribution patterns or temporal trends in seawater pH driven by physical, chemical, and biological processes. For example, using time as a predictor offers the same information for all grids, with insufficient information to capture the regional differences in pH change rates. The key predictors are thus environmental variables related to the physical, chemical, and biological processes that directly affect pH. In some regions where the environmental variables sufficiently reflect the factors influencing pH or where spatial and temporal pH patterns are not notable, adding latitude, longitude, and time as predictors does not contribute sufficient information and can not effectively reduce predicting errors. Instead, it may lead to the overfitting of FFNN, overlooking pH fluctuations within small regions or over short time scales. Therefore, latitude, longitude, and time were excluded from predictors in these areas.

The description of the selective use of sampling location and time information as predictors was added in section as the following :

"Spatial and temporal variables, such as latitude, longitude, and time, are directly related to the spatial or temporal pH patterns rather than the factor driving pH variations. This means these variables are often co-correlated with other input environmental variables. In some regions where the environmental variables sufficiently reflect the factors influencing pH or where spatial and temporal pH patterns are not notable, adding latitude, longitude, and time as predictors does not contribute sufficient information and cannot effectively reduce predicting errors due to the co-correlation with other predictors. In this case, these spatial-temporal variables are not selected as predictors (Tables 2 and 3). "

Broullón, D., et al. (2019). A global monthly climatology of total alkalinity: a neural network approach. Earth System Science Data, **11**(3): 1109-1127.

- Adding a column to Table 1 to indicate which process each variable is associated with would be informative.
Response: Thanks for the suggestion. The related processes have been added in Table 1 as the following:
Thank you for incorporating the suggestion to add the related processes to Table 1. However, I believe further clarification is still needed for some variables. Specifically, variables like PAR, KD, RRS, and Ta/b may also be associated with the biological production of organic matter, as they are crucial in this context. Ensuring that these variables are linked to biological processes in the table will provide a more complete understanding of their roles. I don't think that 'Supplementary for lacking interannual variability of other variables, or potential correlation with unclear process affecting pH' is relevant for

these variables.

Response: Thanks for the suggestion. We have modified the description of the linked processes of used products in Table 1 as the following:

PAR and KD490 are related to the light penetration and availability in aquatic systems influencing phytoplankton photosynthesis;

RRS and Ta/b are related to the phytoplankton composition and suspended particulate matter as indicators of biological productivity.

-Line 191: The paragraph is unclear. The statement, "Therefore, the uncertainty of our pH product was directly estimated from the FFNN pH predicting errors, instead of synthesizing the inherent uncertainty of each used predictor product," needs further clarification. How was this done?

Response: As described in equation (2), the uncertainty was estimated from local pH value and pH predicting error in the corresponding province. For the uncertainty in certain grid, we first convert pH predicting error in the corresponding province into difference of $[H^+]$, by logarithm transfer of predicted and GLODAP measured pH and then calculating RMSE. Subsequently, the RMSE of $[H^+]$ was transferred to pH uncertainty based on the local pH value.

$$\sigma = -\log_{10}(10^{-pH_0} - RMSE_{[H^+]}) - pH_0$$

where RMSE[H+] was the RMSE of $[H^+]$ converted from FFNN pH predicting error in each vertical layer and in each biogeochemical province. pH_0 was the local predicted pH value in the grid that uncertainty was estimated. Due to missing inherent uncertainty of particular predictor product, estimating uncertainty from inherent uncertainty of used predictor products was unfeasible.

Thank you for the detailed response, but the explanation is still unclear regarding how the method described provides local uncertainties. Specifically, the statement *"the uncertainty of our pH product was directly estimated from the FFNN pH predicting errors, instead of synthesizing the inherent uncertainty of each used predictor product"* remains ambiguous. While equation (2) describes converting pH prediction errors into RMSE for $[H^+]$ and then back to pH uncertainty, it's still not clear how this process provides uncertainty estimates at a local (grid-specific) level. Could you provide more detailed clarification on how this method operates for each grid and how the local predicted pH value (pH_0) factors into the uncertainty calculation? Additionally, a more intuitive explanation of why inherent uncertainty from each predictor was not feasible would help clarify this point for readers.

Response: Thanks for the suggestion. Equation (2) is calculated for all grids separately, yielding the local uncertainty for each grid. In the calculation of Equation (2), the pH_0 is the local predicted pH value, and the RMSE is the FFNN error of the province to which the grid belongs, which is used as the local error. The resulting σ represents the local uncertainty for the certain grid. In this approach, since not all grids have observational data to calculate the local error, the local predicted pH value serves to convert the overall province FFNN error into local errors for contained grids by logarithm transfer, allowing for the calculation of local uncertainty, and also play an important role in differentiating the uncertainty among grids with the same error but different pH values.

The reason why it is not feasible to calculate the total uncertainty by combining the inherent uncertainties of different predictor products is that the calculated via the error propagation relies on the partial derivatives of pH to each predictor. However, since the non-linear relationships established by the

FFNN do not have a specific formula, calculating the partial derivatives is extremely difficult. Therefore, it is not possible to use the basic error propagation method to integrate the inherent uncertainties of different products.

The original text has been modified as the following:

"Subsequently, the pH values were shown as pH$_0$±σ at each given pH0 value, and the local uncertainty σ stem from FFNN reconstruction errors was calculated as the following:

$$\sigma = -\log_{10}(10^{-pH_0} - RMSE_{[H^+]}) - pH_0 \tag{2}$$

where RMSE$_{[H^+]}$ was the RMSE of [H$^+$] converted from FFNN pH in each layer of all 14 biogeochemical provinces, pH$_0$ was the local FFNN predicted pH value. The local uncertainty σ calculated by this method is simultaneously related to the pH reconstruction error and local pH level which serves to convert the overall province FFNN error into local errors and better distinguishes the differences in uncertainty across different regions. The uncertainty of products used as pH predictors is one ineluctable source for pH reconstruction errors of the FFNN model. However, the direct estimation of pH uncertainty from summing the uncertainty of each used product is not feasible. Combining the inherent uncertainties of different predictor products via error propagation relies on the partial derivatives of pH to each predictor, but the non-linear relationships established by the FFNN do not have a specific formula, leading to the difficulty in calculating the partial derivatives. Therefore, the local uncertainty of our pH product was directly estimated from the regional FFNN pH reconstruction errors and local pH values following formula (2), instead of synthesizing the inherent uncertainty of each used predictor product through the propagation of errors."

- Moreover, it would be interesting to add comparison against qualified pH data from BGC-Argo dataset.

Response: Thanks for the suggestion. Comparison of global scale trend has been added in Table 4. The BGC ARGO pH data qualified by IMOS has been added in the validation section. Different from the validation results based on the GLODAP dataset, the RMSE between FFNN pH and BGC ARGO pH data is higher in the deep ocean. Only the bias between FFNN pH and BGC ARGO pH data tends to increase with depth in most basins. In contrast, greater biases between FFNN pH and GLODAP pH occur mainly in the surface layer. Especially in the Southern Ocean, the bias between FFNN pH and GLODAP pH is nearly zero below 1000 m, notably lower than biases between FFNN pH and BGC ARGO pH data ranging from 0.053 to 0.076. This may be primarily attributed to the discrepancies between GLODAP dataset and the BGC ARGO dataset in the deep ocean, as our product was based on GLODAP dataset and small biases with GLODAP pH were observed in the deep ocean.

Thank you for incorporating the comparison with the BGC-Argo pH data into the validation section. However, there are several important points that still need to be addressed:

• *BGC-Argo* should be written with proper formatting (i.e., "BGC-Argo," not all caps for "Argo").

Response: Thanks for the suggestion. The formatting has been corrected to "BGC-Argo" in all places.

• It would be valuable to include a reference to the BGC-Argo dataset, such as Claustre et al. (2020) https://www.annualreviews.org/content/journals/10.1146/annurev-marine-010419-010956.

Response: The reference to the BGC-Argo dataset has been added as the following:

Argo. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC). SEANOE, https://doi.org/10.17882/42182, 2024.

Claustre, H., Johnson, K. S., & Takeshita, Y. Observing the global ocean with biogeochemical-Argo. Annual Review of Marine Science, 12(1), 23-48, https://doi.org/10.1146/annurev-marine-010419-010956, 2020.

• As it is mentioned that only data qualified by IMOS were used, does this imply that the validation was limited to the Southern Ocean and data from CSIRO? If so, the validation is not truly global. For a more comprehensive validation, I suggest using data from all DACs (Data Assembly Centers), accessible via the GDACs (Global Data Assembly Centers). There are two GDACs available: US GDAC and France Coriolis GDAC (https://argo.ucsd.edu/data/data-from-gdacs/).

Response: Thanks for the suggestion. The dataset of IMOS-Argo Profiles-biogeochemical data (https://catalogue-imos.aodn.org.au/geonetwork/srv/eng/catalog.search#/metadata/2223b7f2-4bac-4ff1-9b1e-aae9ac58deef) we used in previous revision also includes global BGC-Argo profiles. However, the used data contains both real-time mode and delayed-mode pH-adjusted data. According to the suggestion for robust validation, only delayed-mode data from France Coriolis GDAC has been used now (https://doi.org/10.17882/42182). However, much of the delayed-mode data concentrates on years after 2020, out of the range of our product. The used delayed-mode data is currently mainly distributed in the Southern Ocean (see Figure S6 in the next response). The revised section is as the following:

3.1.2 Validation based on BGC-Argo float pH measurements

Comparison with time series observations in deeper oceans suggested that the distribution of pH reconstruction errors with depth varies notably across different stations. To better assess the performance of FFNN in the reconstruction of pH at different depths, the FFNN reconstructed pH was further evaluated by comparing with independent BGC-Argo delayed-mode pH-adjusted data with quality control flag 1 at various depths (Argo, 2024), with spatial positions showing in Figure S6. Different from the validation results based on the GLODAP dataset, the RMSE between FFNN pH and BGC-Argo pH data in the intermediate layer is 0.051, higher than 0.035 in the mixed layer (Figures 8a and 8b). In both the mixed layer and intermediate layer, most samples were evenly distributed around the y=x line. However, in the intermediate layer, some samples were slightly offset and distributed below the y=x line, which may be the main reason for the notably higher RMSE between FFNN pH and BGC-Argo pH data in the intermediate layer. Overall, there is a good linear correlation between FFNN reconstructed pH and independent BGC-Argo pH data, with R² values of 0.73 and 0.84 in the mixed layer and intermediate layer, respectively.
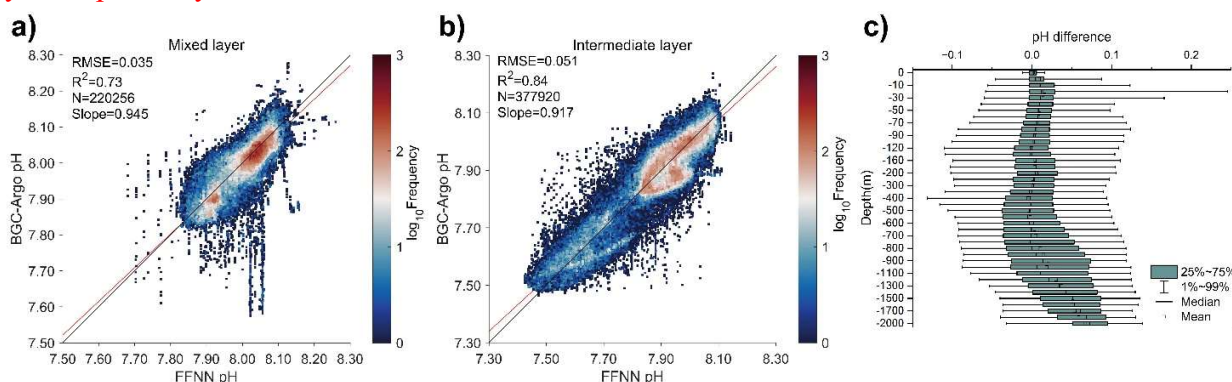
Figure 8. Difference between FFNN pH and BGC-Argo floats pH. a) comparison between FFNN pH and BGC-Argo floats pH in the mixed layer; b) comparison between FFNN pH and BGC-Argo floats pH in the intermediate layer; c) Statistical

distribution of pH difference (FFNN pH minus BGC-Argo floats pH) at different depth levels. FFNN pH: pH data reconstructed in this work; BGC-Argo pH: pH data from BGC-Argo data from France Coriolis GDAC (Argo, 2024).

The distribution of pH differences between FFNN pH and BGC-Argo pH data at different depths reveals relatively smaller biases above 500 m (Figure 8c). However, below 500 m, the bias between FFNN pH and BGC-Argo pH data increases with depth and was the most remarkable at 2000 m. Comparing the pH bias calculated based on BGC-Argo dataset and GLODAP dataset, it is evident that only the bias between FFNN pH and BGC-Argo pH data tends to be more notable in deep areas except the Pacific Ocean (Table 5). In contrast, greater biases between FFNN pH and GLODAP pH occur mainly in the surface layer, with the most in the surface Indian Ocean. This disparity in distribution patterns between biases based on BGC-Argo dataset and GLODAP dataset is most remarkable in the Southern Ocean, where the bias between FFNN pH and GLODAP pH is nearly zero below 1000 m, compared to biases between FFNN pH and BGC-Argo pH data ranging from 0.040 to 0.068. These differences between FFNN pH and BGC-Argo pH data are primarily attributed to the discrepancies between GLODAP dataset and the BGC-Argo dataset in the deep ocean, as our product was based on the GLODAP dataset and small biases with GLODAP pH were observed in the deep ocean.

Table 5. pH bias by area and depth computed with BGC-Argo and GLODAP dataset.

Area		0-50 m	50-200 m	200-500 m	500-1000 m	1000-1500 m	1500-2000 m
Pacific	BGC-Argo	0.028	0.016	-0.003	-0.013	0.027	-0.004
	GLODAP	-0.001	-0.001	0.000	0.000	0.000	-0.001
Atlantic	BGC-Argo	0.018	0.019	0.013	-0.021	0.031	0.068
	GLODAP	0.000	0.000	-0.001	-0.001	0.000	0.000
Indian	BGC-Argo	0.023	0.034	0.025	-0.022	0.000	0.036
	GLODAP	-0.006	-0.001	-0.003	-0.004	-0.004	-0.001
Southern	BGC-Argo	0.008	0.000	0.001	0.015	0.040	0.068
	GLODAP	0.004	0.001	0.001	0.000	0.000	0.000
Global	BGC-Argo	0.012	0.004	0.001	0.008	0.036	0.057
	GLODAP	-0.001	0.000	0.000	-0.001	-0.001	0.000

• A geographical map of BGC-Argo pH profiles should be included to visualize where the validation was performed.

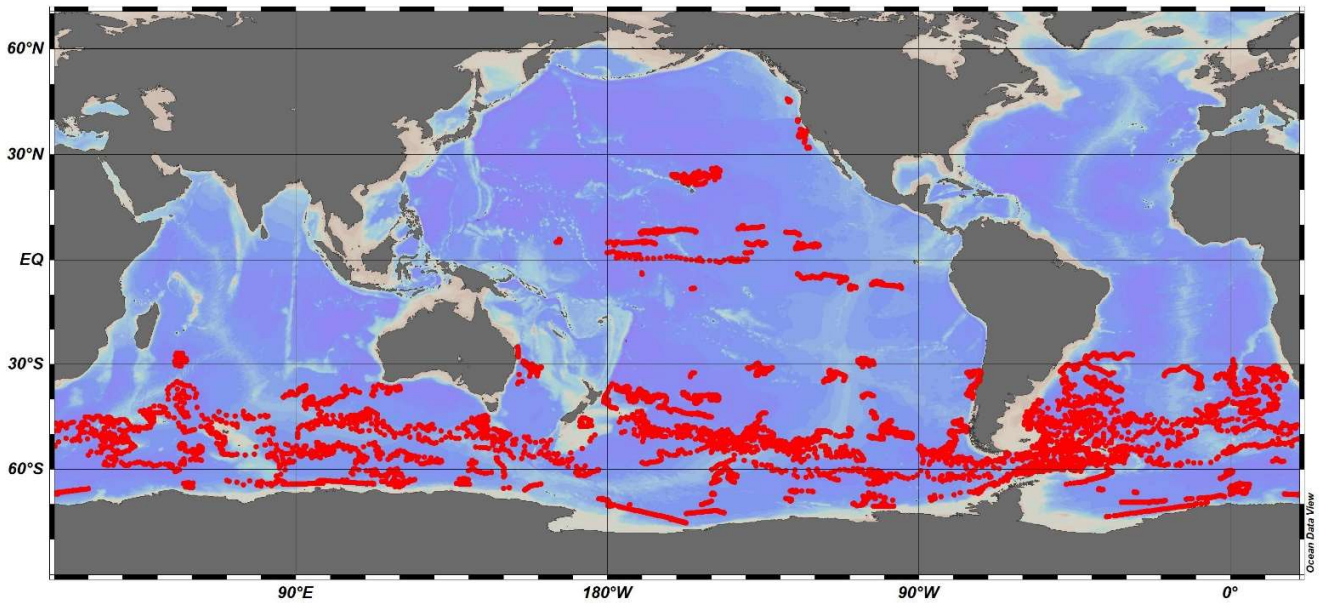Response: The geographical map of BGC-Argo pH profiles has been added in the supplement as the following.

Figure S6. Station map of used delayed-mode BGC-Argo pH-adjusted data with quality control flag 1.

• Specific details regarding the data used in the validation should be provided. Only delayed-mode pH-adjusted data with QC (Quality Control) 1 applied should be used for a robust comparison.

Response: Thanks for the suggestion. The details about used BGC-Argo data were added as the following :

"For better evaluating the performance of FFNN below the surface, the constructed pH product was also compared to independent delayed-mode pH-adjusted data with quality control flag 1 from the biogeochemical-Argo (BGC-Argo) profiles from Global Data Assembly Centre (Claustre et al., 2020; Argo, 2024)."

• Additionally, BGC-Argo should be acknowledged in the acknowledgments, following the guidelines here: https://argo.ucsd.edu/data/acknowledging-argo/.

Response: The BGC-Argo has been acknowledged in the acknowledgments as the following:

"We thank BGC-Argo for sharing the pH float data. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it (http://www.argo.ucsd.edu, http://argo.jcommops.org). The Argo Program is part of the Global Ocean Observing System. "