Reviewer Comment #1

Context and general comment :

The continuous absorption of anthropogenic CO_2 by the ocean leads to ocean acidification, which threatens marine ecosystems. While the acidification rate has been extensively documented at the surface, data for deeper waters remain limited. Zhong et al. address this gap by presenting a comprehensive, monthly, four-dimensional, 1°×1° gridded global seawater pH dataset, covering the years 1992 to 2020 and depths from the surface to 2000 meters.

This dataset was developed using machine learning algorithms trained on pH observations from the Global Ocean Data Analysis Project (GLODAP). The methodology employed is a three-step process: 1) self-organizing map neural network for bioregionalization, 2) a stepwise algorithm for predictors selection, and 3) feed-forward neural networks (FFNN) for non-linear regression. The resulting pH product is a valuable resource for studying subsurface ocean acidification and for validating or initializing biogeochemical models. The product is made publicly available through the Marine Science Data Center of the Chinese Academy of Sciences.

Overall, the article is well-written and the figures are clearly presented.

Despite the significance of this new 3D pH product for the scientific community, the article has some notable shortcomings. There is a lack of details in the methodology section, which makes it challenging to fully evaluate the robustness of the method and comprehend the implications involved.

Response: Thank you very much for agreeing that our data has significant value. The concern about the lack of details in the methodology section is important and helpful for us to improve the manuscript quality. We have revised the manuscript according to the comments, which may provide a clearer understanding of the methods used in this work. The detailed changes were listed in the response to specific comments.

Specific Comments:

Title:

It may be valuable to the reader to add the information that the estimations are depth-resolved, resulting in a 3D product, which is the principal novelty of this methodology.

Response: Thanks for the suggestion. The title has been changed to “A global monthly 3D-field of seawater pH over 3 decades: a machine learning approach”.

Abstract:

- *Lines 15-17:* "Here, we present a monthly four-dimensional 1°×1° gridded product of global seawater pH, derived from a machine learning algorithm trained on pH observations at total scale and in-situ temperature from the Global Ocean Data Analysis Project (GLODAP).": The role of temperature in the methodology is unclear. Even after reading the entire paper, the specific role of temperature compared to other inputs remains ambiguous.

Response: Temperature is an important factor affecting seawater pH, and was used as a pH predictor in almost all regions in this work. In many regions temperature is considered one of the most important predictors and ranks very high in the predictor list in Tables 2 and 3, as the sort order represents the predictor's capacity to reduce predicting errors. Except the relative importance, the role of temperature is the same as other predictors listed in Tables 2 and 3. All selected predictors were treated with same process and input into the neural networks.

The reason for mentioning temperature here is not to show its important role, but to distinguish with the product standardized to a temperature of 25°C. For better clarity, the original text has been modified as the following:

"Here, we present a monthly four-dimensional 1°×1° gridded product of global seawater pH at total scale and in-situ temperature, derived from a machine learning algorithm trained on pH observations from the Global Ocean Data Analysis Project (GLODAP)."

- *Line 18:* I suggest rephrasing the method description for clarity. Consider stating: "A three-step machine learning-based algorithm was used..."

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

"A three-step machine learning-based algorithm was used to construct the pH product, incorporating region division by the self-organizing map neural network, predictor selection by the stepwise regression algorithm that adds and removes variables from network inputs based on their contribution to reducing predicting errors, and non-linear relationship regression by feed-forward neural networks (FFNN)."

- *Line 19:* The term "stepwise" may not be clear to the readers. Consider elaborating or using a more descriptive term.

Response: Thanks for the suggestion. The term "stepwise" has been changed to "the stepwise regression algorithm that adds and removes variables from network inputs based on their contribution to reducing predicting errors" for better clarity.

Introduction:

- The introduction appears to be missing some crucial references. For example, it would be beneficial to acknowledge that the methodology is inspired by the work of Landschützer et al. (2014) and references following; i.e. a SOM-FNN approach. Additionally, it is important to mention that this SOM-FNN approach has already been applied to the 3D reconstruction of DIC by Keppler et al., 2020 (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2020GB006571). These references should be cited to provide a more comprehensive background.

Response: Thanks for correcting me. Landschützer is a scientist whom I highly respect, and I have learned a lot of knowledge from his work. The missing references have been added in the introduction and the text has been modified as the following:

"Recent applications of machine learning methods in the global mapping 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 ($p$$CO_2$, Landschützer et al., 2014; Chau et al., 2022; Zhong et al., 2022), dissolved inorganic carbon (DIC, Broullón et al., 2020; Keppler et al., 2020), and alkalinity (Broullón et al., 2019; Gregor and Gruber, 2021). Additionally, the 3D-field mapping of DIC was also conducted using the SOM-FFNN method, which produced monthly climatological data but lacked interannual variability (Keppler et al., 2020). These methods have inspired our methodology for constructing the global gridded seawater pH dataset. Until now, only surface ocean gridded pH products are available in acidification research, including the 1° JMA product (Iida et al., 2021), the 1° OceanSODA-ETHZ product (Gregor and Gruber, 2021), the 0.25° remote-sensing-based product (Jiang et al., 2022), and the 0.25° CMEMS-LSCE product (Chau et al., 2024), which were derived from mapping $p$$CO_2$, DIC, or alkalinity using machine learning algorithms and subsequently calculating pH with the CO2SYS program (Lewis and Wallace, 1998)."

- L*ine 52 :* where existing pH surface products are listed, there is a lack of references and details. It is crucial to include comprehensive citations of existing products (e.g. LSCE-FFNN), especially those used for comparison in Table 4 (+ the missing ones, see my comment below).

Response: Thanks for the suggestion. The citation of Chau et al., 2024 for the CMEMS-LSCE product and Jiang et al., 2022 for their remote sensing product have been added. The product names, spatial resolutions, and details in method has been added as the following:

"Until now, only surface ocean gridded pH products are available in acidification research, including the 1° JMA product (Iida et al., 2021), the 1° OceanSODA-ETHZ product (Gregor and Gruber, 2021), the 0.25° remote-sensing-based product (Jiang et al., 2022), and the 0.25° CMEMS-LSCE product (Chau et al., 2024), which were derived from mapping pCO_2, DIC, or alkalinity using machine learning algorithms and subsequently calculating pH with the CO2SYS program (Lewis and Wallace, 1998)."

- The paper does not explain why the product spans the period 1992-2020. It would be helpful to provide a rationale for this timeframe and discuss why it does not cover a longer period, both in the past and up to the present (i.e., year-1).

Response: Thanks for the suggestion. The product starts from 1992 because the SSH, MLD, and W velocity of ocean currents used as pH predictors from the ECCO2 cube92 product also start from 1992. Data after 2020 is unavailable because our surface ocean pCO_2 product has not been updated beyond 2020. The pH data will be updated to cover the period 1992-2023 after the update of our pCO_2 product and the NOAA Greenhouse Gas Marine Boundary Layer Reference product (currently updated only until 2022). The explanation for the covered period has been added at the end of the section "2.3 pH product construction" as the following:

"The pH data earlier than 1992 is unavailable because the predictors used from ECCO2 cube92 product (Menemenlis et al., 2008) also start from 1992. Data after 2020 is limited by the coverage of used surface ocean pCO_2 product and will be updated in future works."

- L*ine 55:* The reference to GLODAP by Lauvset et al. (2022) refers to GLODAPv2.2022 and should be cited as such throughout the article (instead of 'GLODAP'). Additionally, if the authors re-run their model during the review process, it is suggested to use the latest version of GLODAP, i.e., GLODAPv2.2023. The updated reference can be found here: https://essd.copernicus.org/articles/16/2047/2024/essd-16-2047-2024-discussion.html

Response: Thanks for the suggestion. The latest version of GLODAP has been used, and the citation has been updated to Lauvset et al., 2023 and added in all places that GLODAP was mentioned.

Methods:

Line 60: It is unclear how temperature is used in the methodology. Additional explanation is needed to clarify its specific role and contribution.

Response: Here we mentioned temperature because the pH reconstructed in this work is at in-situ temperature, not at the commonly used 25°C. Additionally, temperature was used as one of pH predictors and played the same role with other variables technically, as shown in Formula 1 ($\mathrm{pH} = f(\mathrm{Predictors}_1, \mathrm{Predictors}_2, \dots, \mathrm{Predictors}_\mathrm{N})$). Here in-situ temperature was mentioned to avoid the product being considered a 25°C product, as some researchers may convert the pH measurements or pH products to a normalized temperature of 25°C.

Lines 69-73: The inclusion of other indices, such as the Northern Oscillation Index, should be considered.

Response: Thanks for the suggestion. We believe that parameters like the Northern Oscillation Index have smaller impact on pH than currently used indices, and are not considered here. However, we will evaluate the impact of the Northern Oscillation Index on pH in future work. These indices will be added in the future works to be tested if they can contribute to reduction of pH predicting errors.

Lines 71-73: The LSCE-FFNN product provides total alkalinity and DIC data monthly from 1985 to year-1: data available from the Copernicus Marine Service (https://data.marine.copernicus.eu/product/MULTIOBS_GLO_BIO_CARBON_SURFACE_REP_015_008/description). If authors refer to 3D products, then it has to be clearly mentioned. For 3D estimations, DIC is available monthly from 2004 to 2019 from MOBO-DIC (Keppler et al., 2020).

Response: Thanks for the suggestion. The description was about 3D products of DIC and alkalinity, and has been corrected as the following:

"However, 3D field products with sufficient time and spatial coverage are currently not available for these two variables, so climatological 3D products were used for better pH spatial distribution."

Section 2.2 :

- *Lines 103-105:* The rationale for using these specific parameters to define bioregions needs clarification, especially if they are not significant in the stepwise algorithm for determining important parameters in relation to pH.

Response: Thanks for the suggestion. The parameters used to define bioregions are related to physical and biological processes affecting pH, with most being selected as predictors in many bioregions, except for chlorophyll concentration. However, this does not imply that chlorophyll concentration is unrelated to pH changes. Chlorophyll concentration is highly related to photosynthesis, which affects pH by influencing pCO_2. Its role as a pH predictor is replaced by pCO_2. Currently, the uncertainty of pCO_2 products is higher than that of satellite remote sensing chlorophyll products, and using different pCO_2 products may affect the dividing of bioregions. Although the current bioregions are not perfect, we have set a small number of bioregions with broad coverage to reduce the impact of potential inaccuracies in bioregions division. In future work, we will further improve the parameters used to define bioregions based on the currently selected predictors.

- *Line 106 :* The criteria and process for merging provinces with fewer than ten connected grids or less than 100 GLODAP pH measurements should be rephrased and/or detailed because it is unclear.

Response: Thanks for the suggestion. The SOM may accidently generate three types of small "island" provinces: provinces consisting of many disconnected grids across different regions, provinces covering very small areas with fewer than 10 connected grids, and provinces with an insufficient number of GLODAP pH measurements. These provinces are not helpful in the pH product construction, as the pH predicting errors tend to be notably higher due to the extremely few training samples for FFNNs. A similar process can be also found in previous SOM-based research, such as Landschützer et al. (2014), which removed small "island" provinces with a surface area smaller than 10 connected grid cells. The origin text has been modified as the following: "Subsequently, the small "island" provinces with fewer than ten connected grids or covered by fewer than 100 GLODAP pH measurements were merged with the nearest neighboring provinces, as the pH predicting errors tend to be notably higher due to the extremely few training samples in the non-linear relationship fitting by networks."

- *Line 107 :* The need for manual subdivision of provinces separated by continents requires further explanation. Why not using same bioregion even if it is not in the same ocean, it is possible that the underlying processes are equivalent and so that the FFNN will be performant in both basin?

Response: Thanks for the suggestion. The physical process may also affect seawater pH distribution and are different between basins separated by continents, such as surface ocean currents. However, these physical mixing and transport processes are actually separated by lands, so we manually divided provinces separated by continents.

- *Lines 112-114 :* The sentence on the division of ocean areas into different layers also requires further details on which drivers are important for each layer as it is stated that drivers differ depending on the layers. Moreover, following this statement, why using the same bioregions for deeper layers?

Response: In the mixed layer, seawater pH is notably influenced by the seasonal cycle of environmental conditions and CO_2 exchange between the surface ocean and atmosphere. In the intermediate layer, seawater pH experiences much weaker seasonal changes and is largely affected by the biological drawdown of organic matter. Such differences in pH drivers can also be observed in the selected predictors in Tables 2 and 3, with the most important predictor in many provinces being surface ocean pCO$_2$ in the mixed layer and phosphate concentration in the intermediate layer.

The variables used for division of regions are fully available in the surface ocean, but some variables lack information at different depth, making it difficult to divide regions separately in the deep ocean. Additionally, to maintain consistency in geographic regions between the two vertical layers with existing vertical seawater mixing, it is unnecessary to divide regions separately in the mixed and intermediate layers as they are not completely separate areas. Therefore, we applied the surface biogeochemical provinces to the deeper ocean as well.

Section 2.3 and Table 1 :

- The choice of a single-layer FFNN instead of a multi-layer network should be justified. Has this been tested ?

Response: The FFNN with single hidden layer has a smaller scale with faster training and calculating speed, and is also convenient for adjusting the number of neurons. We also compared different neural network structures. With a similar number of neurons, the impact of different structures on error is relatively small. By increasing the number of neurons, a fitting capability comparable to that of multi-layer neural networks can be achieved. Additionally, in previous studies reconstructing surface ocean pCO$_2$ gridded data, we also used single-layer neural networks and verified their sufficient fitting capability to reconstruct global ocean gridded product of carbonate system variables.

- 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.

Denvil-Sommer, A., et al. (2019). "LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean pCO_2 over the global ocean." Geoscientific Model Development 12(5): 2091-2105.

- 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."

- The choice of the ECCO2cube92 model should be discussed and better supported by citations in the text.

Response: Thanks for the suggestion. We used the ECCO2 cube92 product due to its wide temporal and spatial coverage, which starts from 1992 and is continuously updating to the present. Moreover, the MLD and SSH data from ECCO2 were also used in previous research on mapping of ocean carbonate system variables and proved the reliability, for example, Landschützer, et al. (2014) and Chau, et al. (2024). A description has been added in "2.1 Data sources and processing" section as the following:

"Products used for variables listed in Table 1 was chose due to their sufficient temporal and spatial coverage and the application in previous research on mapping carbonate system variables. For example, the ECCO2 MLD product has been used in reconstruction of the CMEMS-LSCE surface ocean carbonate system variables product (Chau, et al., 2024) and the MPI-SOM-FFN pCO2 product (Landschützer et al., 2014)."

Landschützer, et al. (2014). Recent variability of the global ocean carbon sink. Global Biogeochemical Cycles 28(9): 927-949.

Chau, et al. (2024). CMEMS-LSCE: a global, 0.25°, monthly reconstruction of the surface ocean carbonate system. Earth System Science Data 16(1): 121-160.

- MEI should be defined as the Multivariate ENSO Index.

Response: Thanks for the suggestion. The name of MEI has been corrected.

- 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:

Table 1. Data products used as pH predictors.

Predictor	Abbreviation	Data product and reference	Resolution	Related process to affect pH
Sine of (latitude · π/180°)	sin(Lat)	-	-	Sample position and time of GLODAP pH measurements
Sine of (longitude · π/180°)	sin(Lon)	-	-	
Cosine of (longitude · π/180°)	cos(Lon)	-	-	
Number of months since January 1992	N_{mon}	-	-	
Year	Year	-	-	
Month	Month	-	-	
Depth	Depth	-	-	
Temperature and monthly anomaly	Temp, $Temp_{anom}$	IAP global ocean temperature gridded product (Cheng et al., 2016; 2017)	1°, monthly since 1940, 0-2000 m with 41 levels	State of carbonate system
Salinity and monthly anomaly	Sal, Sal_{anom}	IAP global ocean salinity gridded product (Cheng et al., 2020)	1°, monthly since 1940, 0-2000 m with 41 levels	
Climatological total alkalinity	Alk	AT_NNGv2_climatology (Broullón et al., 2019)	1°, monthly climatological, 0-5500 m with 102 levels	
Climatological dissolved inorganic carbon	DIC	TCO2_NNGv2LDEO_climatology (Broullón et al., 2020)	1°, monthly climatological, 0-5500 m with 102 levels	

Climatological dissolved oxygen	DO	WOA18 (Garcia et al., 2020a)	1°, monthly climatological, 0-5500 m with 102 levels	Biological production and drawdown of organic matter
Climatological nitrate	Nitrate	WOA18 (Garcia et al., 2020b)	1°, monthly climatological, 0-5500 m with 102 levels	
Climatological phosphate	Phosphate			
Climatological silicate	Silicate			
Mixed layer depth and monthly anomaly	MLD, MLD_{anom}	ECCO2 cube92 (Menemenlis et al., 2008)	0.25°, monthly since 1992	Physical mixing of seawater and stratification
Sea surface height and monthly anomaly	SSH, SSH_{anom}			ocean wave, tides, current, and sea-level rise
W velocity of ocean currents at 5 m, 65m, 105m, 195m, and in-situ depth	W_{vel}(5m)−W_{vel}(in-situ)			Ocean current and upwelling
Sea level pressure	SLP	ERA5 (Hersbach et al., 2020)	1°, monthly since 1940	CO_2 exchange between surface seawater and atmosphere
Surface pressure	P_{surf}			
dry air mixing ratio of atmospheric CO_2 and monthly anomaly	xCO_2, $xCO_{2\ anom}$	NOAA Greenhouse Gas Marine Boundary Layer Reference (Lan et al., 2023)	0.25°, weekly since 1979	
Multivariate ENSO Index	MEI	bi-monthly Multivariate El Niño/Southern Oscillation index (Wolter et al., 2011)	monthly since 1979	El Niño and Southern Oscillation
Arctic Oscillation index	AOI	Climate Prediction Center Daily Arctic Oscillation Index (CPC, 2002)	monthly since 1950	Arctic Oscillation
Southern Oscillation Index	SOI	Climate Prediction Center Southern Oscillation Index (CPC, 2005)	monthly since 1951	Southern Oscillation
Bathymetry	Bathy	GEBCO_2022 Grid (GEBCO, 2022)	15 arc-second	Vertical volume of seawater
10 m Wind speed and monthly anomaly	Wind, $Wind_{anom}$	ERA5 (Hersbach et al., 2020)	1°, monthly since 1940	CO_2 exchange between surface seawater and atmosphere
Surface ocean pCO_2	pCO_2	Stepwise FFNN (Zhong et al., 2022)	1°, monthly since 1992	
Climatology of Surface Ocean pCO_2	$pCO_{2\ clim}$	MPI-ULB-SOM_FFN_clim (Landschützer et al., 2020)	0.25°, monthly climatological	
Chlorophyll and monthly anomaly*	Chl, $Chl_{\ anom}$	MODIS-Aqua Chlorophyll Data (NASA, 2022a)	9km, monthly since 2002	Biological production of organic matter
Photosynthetically Available Radiation	PAR	MODIS-Aqua Photosynthetically Available Radiation Data (NASA, 2022b)		
Diffuse attenuation coefficient at 490 nm	KD490	MODIS-Aqua Downwelling Diffuse Attenuation Coefficient Data (NASA, 2022c)		Supplementary for lacking interannual variability of other variables, or potential correlation with unclear process affecting pH
Remote sensing reflectance at 412-678 nm**	RRS412−RRS678	MODIS-Aqua Remote-Sensing Reflectance Data (NASA, 2022d)		
Total absorption at 412-678 nm	Ta412−Ta678	MODIS-Aqua Inherent Optical Properties Data (NASA, 2022e)		
Total backscattering at 412-678 nm	Tb412−Tb678	MODIS-Aqua Inherent Optical Properties Data (NASA, 2022e)		

(*: products from Chlorophyll to Total backscattering are satellite remote sensing products;

**: Remote sensing reflectance, total absorption, and total backscattering both include 10 wavelengths: 412nm, 443nm, 469nm, 488nm, 531nm, 547nm, 555nm, 645nm, 667nm, and 678nm, with each wave length regard as one individual parameter.)

- Consider using merged satellite ocean color data products like OC-CCI or GlobColour for longer time series would help for future usability and sustainability.

Response: Thanks for the suggestion. These products will be used when we update our product.

- Provide details on how the most informative parameters were chosen and how hyperparameters (architecture, number of neurons) were handled in this stepwise process.

Response: Thanks for the suggestion. We have revised the predictor selection section, and added the specific procedure of stepwise FFNN algorithm in Figure 3 as the following:

"(1) Selection of seawater pH predictors in each province using the Stepwise FFNN algorithm (referred as (1) Stepwise FFNN in Figure 3). All the collected products were input into the Stepwise FFNN algorithm to identify the predictors that yield the lowest predicting errors for seawater pH (Zhong et al., 2022). The variation in standard deviation (MAE) calculated by the K-fold cross validation method will feed back to update the input products. The input variables are selected as pH predictors one by one in the way MAE decreases the fastest. Specifically, by comparing predicting errors of using each collected environmental variable in Table 1 as the only predictor input to the FFNN, the variable with the lowest error is selected as the first pH predictor and moved out from the environmental variables list used in the subsequent steps. Subsequently, while keeping the first predictor unchanged, compare predicting errors when using each remaining environmental variable as the second input for the FFNN. The variable with the lowest error is determined to be the second pH predictor. In the same way, new predictors are sequentially determined. This selection process continued through multiple iterations until no further reduction in MAE was observed, regardless of whether a variable was added or removed. The variables identified in previous iterations were then output as the optimal pH predictors. Since both overfitting caused by co-correlation and underfitting caused by an insufficient number of predictors result in significant increases in pH predicting errors, the lowest predicting error is considered to occur between these two states. In order to eliminate potential co-correlation and prevent overfitting, whenever a new predictor is identified, the algorithm also tests whether the predicting error will decrease when sequentially removing each determined predictor. The algorithm individually removes each previously identified predictors immediately after adding one variable as a predictor. If the error decreases after removing a previously determined predictor, this predictor is highly correlated with other identified predictors. Therefore, most of the co-correlation among the selected predictors has been removed in this Stepwise FFNN selection procedure. If products with co-correlations are still selected, some products may provide important additional information in specific regions, leading to a greater reduction in predicting errors compared to the increase caused by overfitting. In each province, pH predictors were selected separately for the mixed layer (Table 2) and intermediate layer (Table 3). In certain polar areas and prior to August 2002 when satellite remote sensing products (products from Zeu to Tb678 in Table 1) were not available, the additional selection of predictors was carried out without the use of satellite remote sensing products (Table S1). These satellite products were not used in the intermediate layer due to low correlation with seawater pH, with no need for additional selection.

All FFNNs used in these two steps have the same structure with a single hidden layer, as using deeper structures tends to cause overfitting and increase pH predicting errors. The number of neurons was determined by comparing predicting errors of FFNNs with different neurons based on the same training samples, testing samples, and pH predictors, and then adopting the number with the lowest predicting error. Specifically, for the stepwise FFNN regression step, the number of neurons in FFNNs was determined using provisional predictors from preliminary experiments with the number of neurons set to 25."

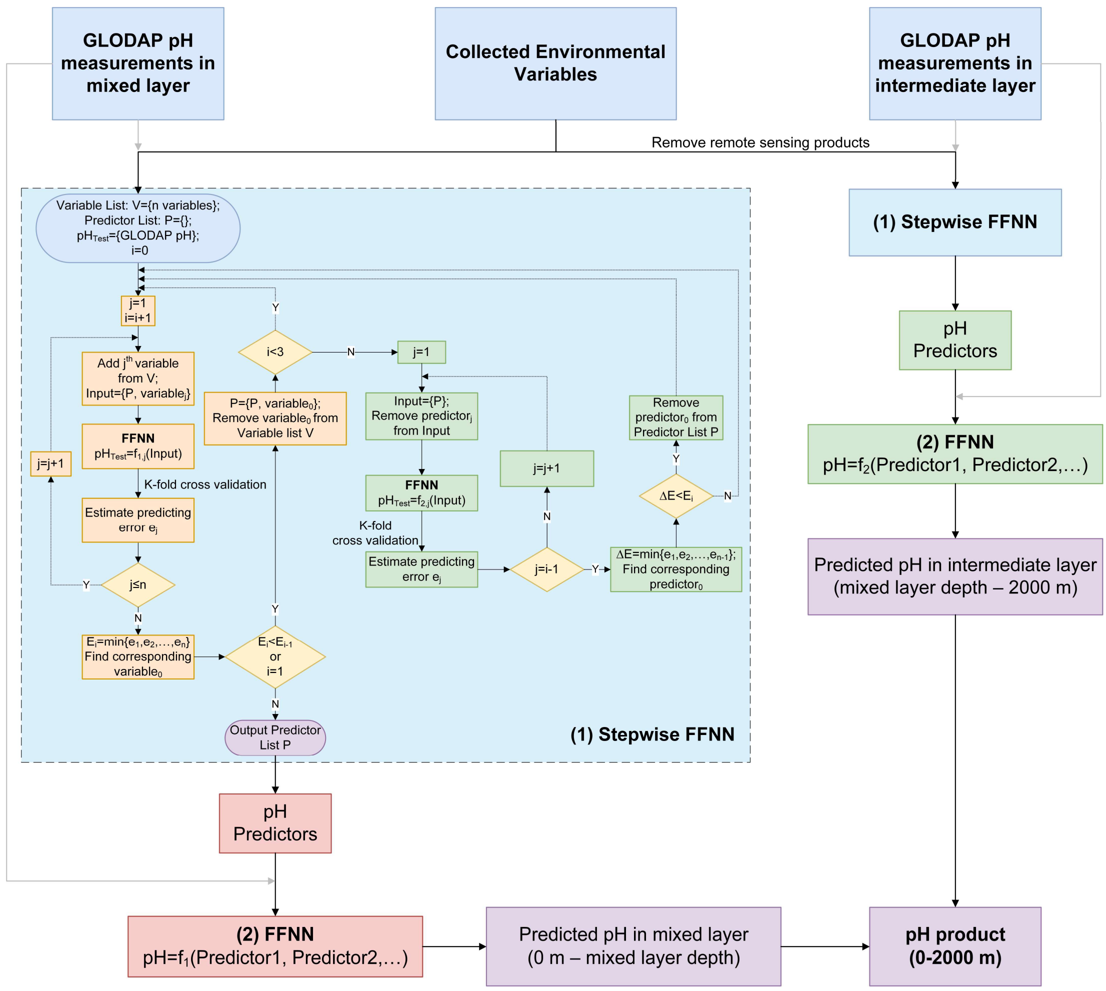

Figure 3: The procedure of pH product construction. (1) Stepwise FFNN: the algorithm for selecting predictors (Zhong et al., 2022); (2) FFNN: fitting the non-linear relationship between seawater pH and its predictors. Collected Environmental Variables: collected products listed in Table 1. pH predictors: the selected most informative variables listed in Tables 2 and 3. Remote sensing products: variables from Chlorophyll to Total backscattering in Table 1. Mixed layer: from 0 m to mixed layer depth; intermediate layer: from mixed layer depth to 2000 m.

- Clarify how co-correlation among selected predictors was removed in the stepwise FFNN selection procedure.

Response: Since the variables with co-correlations provide similar information, the predicting error using two co-correlated predictors will be higher than that using two predictors related to different ocean processes. Therefore, the variable correlated to existing predictors tends to fail to compete with other variables in the predictor selection. Moreover, whenever after a new predictor is identified, the stepwise FFNN algorithm also tests whether the predicting error will decrease when sequentially removing each determined predictor. If a certain predictor is highly correlated with existing predictors, this predictor will be generally removed for to reduce in predicting errors.

- The sentence regarding additional FFNNs trained with predictors in Table S1 for polar areas and periods before August 2002 needs clarification.

Response: The sentence has been modified as the following:

"Since the satellite remote sensing products used in this work lack data during the period before August 2002 and in certain polar areas during winter, the FFNN generated missing values in these grids when remote sensing products were used as predictors. To address these missing values, we selected additional groups of predictors after removing remote sensing products (Table S1), and then trained additional FFNNs to predict pH in grids with missing values. This procedure was the same as the reconstruction process in the intermediate layer, in which the remote sensing products were also not used."

- Discuss Tables 2 and 3 scientifically in the Results section to highlight important processes driving pH variability.

Response: Thanks for the suggestion. We have added a discussion about processes driving pH variability in the end of section 3.2.1 Spatial pH distribution as the following:

"Based on the pH predictors selected by the Stepwise FFNN algorithm, differences in processes driving pH variability were identified between the mixed layer and intermediate layer in most provinces. In the mixed layer, surface ocean pCO_2 was identified as the most informative predictor in many provinces, followed by temperature and nutrient concentration. This suggests that the CO_2 exchange between surface ocean and atmosphere is the primary driver of pH variability, followed by biological CO_2 utilization and seasonal changes in seawater temperature. In contrast, phosphate was identified as the most informative predictor in the intermediate layer, followed by temperature and depth. This suggests that the primary process driving pH variability is the remineralization of organic matter, converting organic carbon into inorganic forms and also releasing nitrogen and phosphorus. Given the notably smaller seasonal temperature changes in the intermediate layer compared to the mixed layer, the selection of temperature as an important pH predictor may indicate a notable influence of ocean warming on seawater pH variability. Additionally, depth was also selected as an important predictor in the intermediate layer. The observed pattern of seawater pH decreasing with increasing depth in most provinces, as suggested by the constructed pH product, may be the main reason."

- Figure 3 is extremely difficult to understand and should be clarified or redesigned.

Response: Thanks for the suggestion. We have revised Figure 3 as the following:

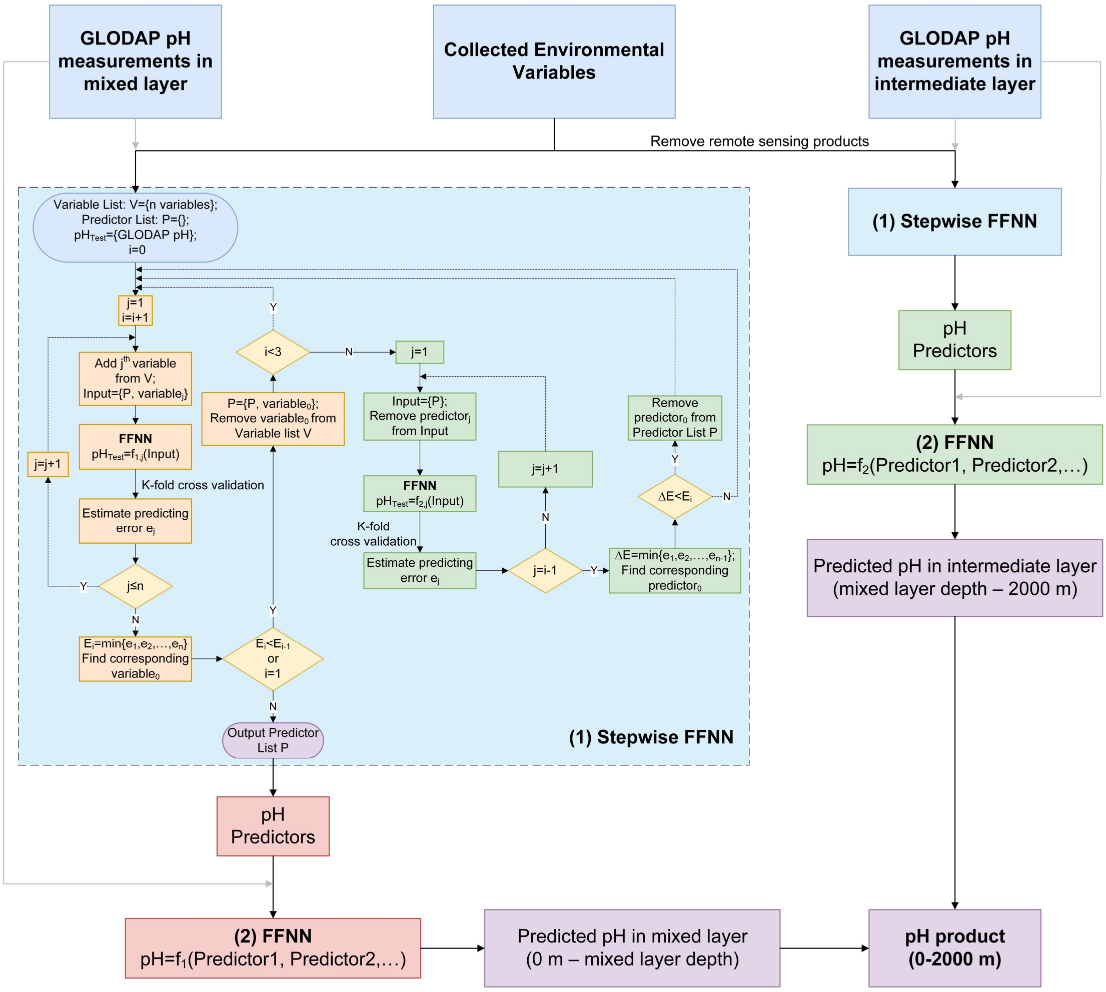

Figure 3: The procedure of pH product construction. (1) Stepwise FFNN: the algorithm for selecting predictors (Zhong et al., 2022); (2) FFNN: fitting the non-linear relationship between seawater pH and its predictors. Collected Environmental variables: collected products listed in Table 1. pH predictors: the selected most informative variables listed in Tables 2 and 3. Remote sensing products: variables from Chlorophyll to Total backscattering in Table 1. Mixed layer: from 0 m to mixed layer depth; intermediate layer: from mixed layer depth to 2000 m.

More generally, the section 2.3 is currently unclear and needs to be rewritten with more detailed explanations.

Response: Thanks for the suggestion. We have revised the section 2.3 and added more details about the method. The specific changes can be found in above response.

Section 2.4:

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.

$$\sigma = -\log_{10}(10^{-\mathrm{pH}_0} - RMSE_{[\mathrm{H}^+]}) - \mathrm{pH}_0$$

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.

Section 3.1:

- *Line 214:* This interpretation might be overstated. The broader value range likely contributes to a better model fit, and pH values exhibit less variability at depth.

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

"The minor difference between the predicting value and the pH measurements and the R2 of 0.97 in the intermediate layer may be caused by less pH variability at depth and better model fit with broader pH value range."

- *Figure 4:* Authors might add the slopes of the linear regression to the statistics.

Response: The slopes and linear regression lines have been added.

- *Line 235:* The impact of the Oxygen Minimum Zone (OMZ) on the product should be discussed more in details.

Response: Although dissolved oxygen was considered the most informative predictor in the Indian Ocean, statistical differences in pH predicting errors were not observed between the OMZ and other areas. We have added a description about the impact of the OMZ on the quality of pH product. The impact of the OMZ on the pH variability will be investigated in future work, as much more efforts are needed and the currently used dissolved oxygen product is only monthly climatological.

- *Figure 5:* This figure requires re-arrangement. The map should be larger, and pH differences against depth should be plotted with depth as the y-axis, as is more common for reading profiles. Additionally, including seasonal variability for each major basin along with yearly variability would be beneficial.

Response: Thanks for the suggestion. We have re-arrangement Figure 5. It is difficult to compare seasonal variability for each major basin, as the seasonal variability of the north and south hemispheres in the Pacific and Atlantic Ocean cancel each other out.

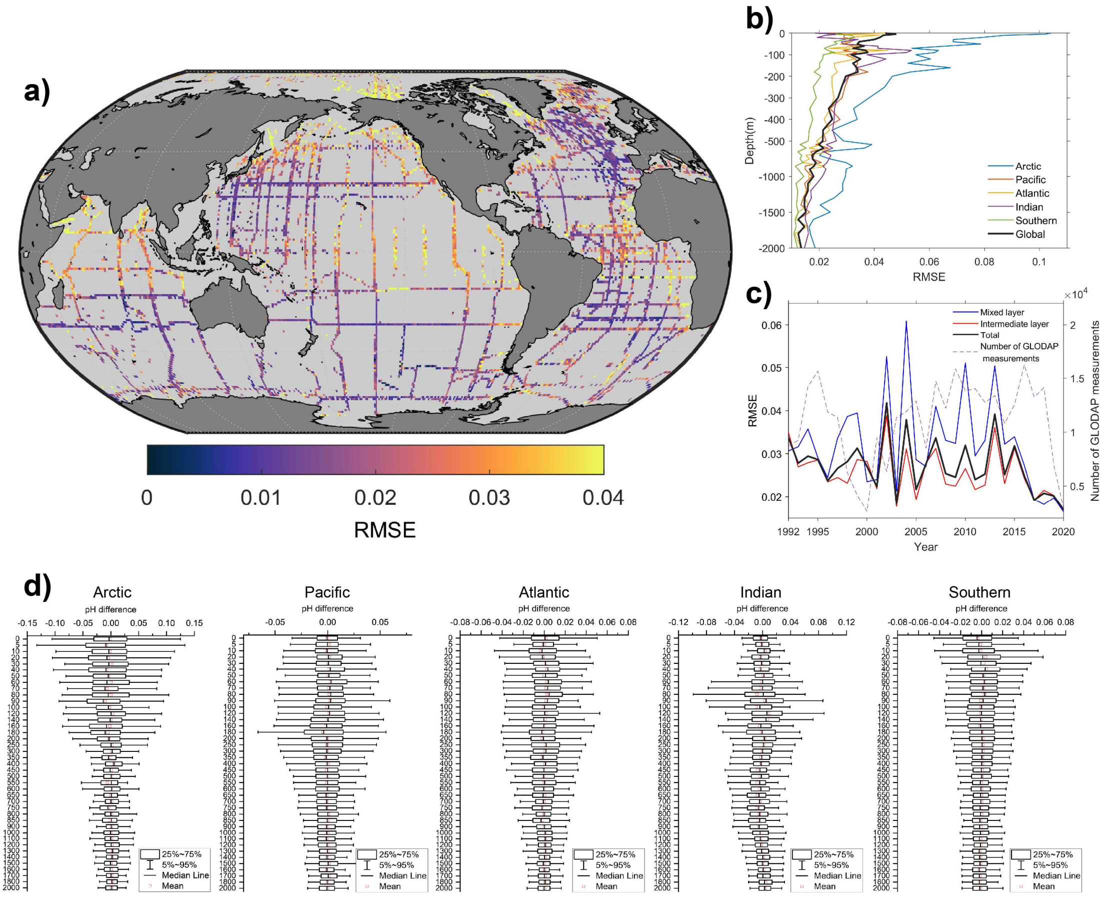

Figure 5: Distribution of RMSE between FFNN predicted pH values and GLODAP pH measurements. a): global spatial distribution of RMSE between FFNN predicted pH and GLODAP pH measurements at 0-2000 m; b): basin average RMSE at different depth; c): temporal distribution of global RMSE; d): Statistical distribution of pH difference between predicted pH values and GLODAP pH measurements in each basin.

- In the validation section, it would be valuable to compare the global scale trend with the Copernicus Marine Service data: https://marine.copernicus.eu/access-data/ocean-monitoring-indicators/globalocean-acidification-mean-sea-water-ph-time-series.

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.

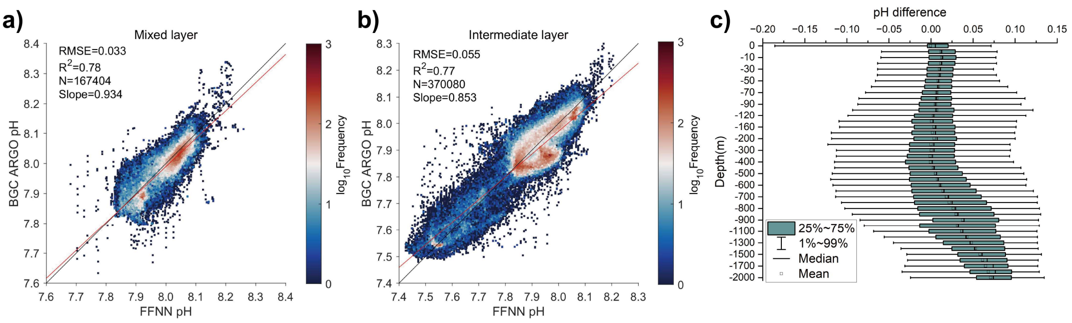

Figure 8. Difference between FFNN pH and BGC ARGO floats pH. a) comparison between FFNN pH and BGC ARGO floats pH in the mixed layer; b) comparison between FFNN pH and BGC ARGO floats pH in the intermediate layer; c) Statistical distribution of pH difference (FFNN pH minus BGC ARGO floats pH) at different depth levels. FFNN pH: pH data reconstructed in this work; BGC ARGO floats pH: pH data from IMOS Biogeochemical ARGO floats core data collection (IMOS 2002-2020, 2021).

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
Arctic	BGC ARGO	-0.006	0.021	0.014	0.026	0.037	0.029
	GLODAP	-0.005	-0.003	0.001	0.000	0.002	0.003
Pacific	BGC ARGO	0.012	0.011	-0.008	-0.015	0.008	0.000
	GLODAP	-0.001	-0.001	0.000	0.000	0.000	-0.001
Atlantic	BGC ARGO	0.016	0.017	0.010	-0.024	0.029	0.068
	GLODAP	0.000	0.000	-0.001	-0.001	0.000	0.000
Indian	BGC ARGO	0.024	0.026	0.014	-0.056	0.002	0.059
	GLODAP	-0.006	-0.001	-0.003	-0.004	-0.004	-0.001
Southern	BGC ARGO	0.011	0.002	0.002	0.027	0.053	0.076
	GLODAP	0.004	0.001	0.001	0.000	0.000	0.000
Global	BGC ARGO	0.011	0.005	0.001	0.021	0.046	0.066
	GLODAP	-0.001	0.000	0.000	-0.001	-0.001	0.000

- *Figure 6 + text:* Comparing to other available pH time series would be interesting. These are listed in the recent ESSD paper by Lange et al. (2024): https://essd.copernicus.org/articles/16/1901/2024/. For instance, the Mediterranean Sea, where data from GLODAP V2 are very scarce, could be validated against the Dyfamed pH time series.

Response: Thanks for the suggestion. We only compared with three station due to their sufficient availability in temporal coverage. The collection by Lange et al. (2024) includes many stations and are helpful for further evaluating our product, but the pH data of particular stations are only available for several years. Therefore, we only added the Iceland Sea, the Irminger Sea, and the DYFAMED station data in Table 4 as the following:

Table 4: Comparison of surface acidification rate with previous product in different time series stations and on a global scale.

Stations	Period	Time series observation	Stepwise FFNN (This study)	JMA (Iida et al., 2021)	CMEMS (Chau et al., 2024)	OS-ETHZ (Gregor et al., 2021)	Copernicus (Copernicus Marine Service, 2020)
BAT	1992~2020	-0.0018 ± 0.0001	-0.0017 ± 0.0007	-0.0018 ± 0.0002	-0.0018 ± 0.0002	-0.0018 ± 0.0002	-
ESTOC	1995~2010	-0.0016 ± 0.0001	-0.0014 ± 0.0005	-0.0022 ± 0.0003	-0.0020 ± 0.0002	-0.0017 ± 0.0003	-
HOT	1992~2020	-0.0018 ± 0.0001	-0.0018 ± 0.0004	-0.0020 ± 0.0001	-0.0021 ± 0.0001	-0.0019 ± 0.0001	-
Iceland Sea	1992~2019	-0.0020 ± 0.0004	-0.0028 ± 0.0002	-0.0030± 0.0003	-0.0015 ± 0.0002	-0.0020 ± 0.0002	-
Irminger Sea	1992~2019	-0.0025 ± 0.0004	-0.0022 ± 0.0002	-0.0027 ± 0.0002	-0.0017 ± 0.0003	-0.0016 ± 0.0003	
DYFAMED	1998~2017	-0.0010 ± 0.0008	-0.0005 ± 0.0003	-	-0.0017 ± 0.0003	-0.0023 ± 0.0004	
Global	1992~2020	-	-0.0015 ± 0.0002	-0.0018 ± 0.0000	-0.0017 ± 0.0004	-0.0018 ± 0.0000	-0.0017 ±0.0002

- *Figure 6:* Discuss the extreme values not reconstructed by the FFNN in the text.

Response: The extreme low values not reconstructed by the FFNN are mainly observed at the BAT station near 2010 and at the HOT station near 2000, under the influence of La Niña events. The extreme high values are mainly observed at the HOT station before 2000, under the influence of El Niño events. Differently, the extreme values not reconstructed by the FFNN are less observed at the ESTOC station, where the surface pH did not notably fluctuate during El Niño/La Niña events. It can be inferred that the extreme values not reconstructed by the FFNN may be due to its underestimating of the impact of El Niño/La Niña events on pH of certain temperate areas.

- *Line 254 :* Chau et al. (2022) may not be the best reference, as they are also model (ML)-based.

Response: The reference has been corrected to González-Dávila et al. (2010).

- *Line 261* : Describe in what specific ways the product differs from other products.

Response: The long-term pH trend of our product at the ESTOC station was only -0.0014 yr^{-1}, notably slower than the trend from -0.0017 yr^{-1} to -0.0022 yr^{-1} of other gridded products, but is still close to the -0.0016 ± 0.0001 yr^{-1} of real observations.

- *Table 4:* More products could be compared, such as Jiang et al. (2022): Remote Sensing of Global Sea Surface pH Based on Massive Underway Data and Machine Learning (https://doi.org/10.3390/rs14102366). Additionally, some products compared here have not been previously cited in the article (refer to the comment on the introduction). The effect of the different time ranges of the different products on the computation of trends should also be analyzed and discussed.

Response: Thanks for the suggestion. The comparison of surface pH trend in time series stations and on a global ocean scale between different products was based on the same period, which all contain years before 2004 not covered by product of Jiang et al. (2022). Therefore, the product of Jiang et al. (2022) was not used, but we added the Copernicus Marine Service product for comparison instead, which has a longer time coverage. The citations of products have been added in the introduction. The pH trends from other products for comparison were all re-calculated using data during same periods, and the description has been added in table remarks.

Section 3.1.2 and Figure 7: Not sure whether this paragraph and figure are necessary.

Response: These contents has been moved to supplement.

Section 3.2:

- *Lines 301-304:* This issue is problematic and should be discussed in more details for the user. Additionally, the significant differences between the GLODAP climatology and this product at 1000 m in the Southern Ocean should be discussed/addressed.

Response: The description has been corrected to focus on the temperate Pacific as the following:

"In the temperate Pacific Ocean, differences in surface pH levels were observed between the west and east in both our product and GLODAP pH climatology, which may be caused by the spread of eastern equatorial seawater with extremely low pH. At the deeper depth of 1000 m, the spatial distribution pattern of FFNN pH product is generally consistent with the GLODAP climatology, despite still some disturbance of bad FFNN performance along the SOM province boundary and the higher FFNN pH in the Southern Ocean. However, the distribution of higher FFNN pH in the region between 35°S and 50°S is consistent with the lower DIC reconstructed by Broullon., et al. (2020)."

The difference in the Southern Ocean may be because the pH observations are sparse and uneven in time and space in the deep Southern Ocean, leading to some pH distribution differences in local areas or depths. However, in the high-latitude regions of the Southern Ocean, our constructed data is also in good agreement with the GLODAP climatology, with an average pH of around 7.9. In the region between 35°S and 50°S, the distribution of higher FFNN pH is consistent with the lower DIC reconstructed by Broullon, et al. (2020).

- *Figure 9:* The longitude of the zonal average should be specified in the caption and/or the text.

Response: The pH values shown at each latitude were averaged from pH values across all longitudes within each major basin. The description has been added in the caption.

-*Section 3.2.2:* The discontinuity problem requires more discussion, both methodologically (explaining why this issue occurs despite the use of the cross-boundary method) and in terms of implications for users. If local uncertainties are available, they should be included in the NetCDFs.

Response: Thanks for the suggestion. The further discussion has been added in the Section 3.2.2. The local uncertainties have been added in the NetCDF files. Here are the added text:

"Although we have used a cross-boundary method to improve the FFNN performance near the SOM and vertical boundary, there are still some discontinuity problems and relatively higher uncertainty. This is because the pH values on two sides of the SOM boundary were reconstructed from two different FFNN models, which were trained with different samples and used different predictors. If one of the FFNN models experienced a worse performance due to insufficient training samples or predictors, the pH values on two sides of the SOM boundary will still differ notably, that is, discontinuity along the boundary. Therefore, the analyze on a regional scale based on pH values near SOM boundaries should be more cautious when using our product."

Section 5:

Authors should provide more concrete examples of applications for their product in the Conclusion.

Response: The description of applications has been added in the Section 5 as the following:

"This product serves as a reference for guiding acidification surveys by providing a general understanding of acidification process at different depths on a basin scale and indicating areas with potential fast or slow acidification rates. Additionally, the pH product brings insights into acidification research and can be used to analyze the influence of specific ocean processes on acidification rates and the broader impacts of acidification on a large scale when direct observations are unavailable."

Typos :

- *Line 78:* was converted to a 1°×1° resolution by averaging **16** 0.25° grids into one 1° grid

- *Line 84:* **(***

- *Line 116:* Therefor**e**

Response: These typos have been corrected.

Data:

I encountered an error when attempting to open the NetCDF file using R. The error message was as follows:

Dans nc_open("/home/user/Data/2012.nc") :

WARNING file /home/user/Data/2012.nc is not compliant netCDF; variable pH is numeric but has a character-type missing value! This is an error! Compensating, but you should fix the file!

Although I didn't receive any warnings when using xarray with Python, this issue should be addressed to ensure compatibility with other tools. Additionally, when opening the dataset with xarray and attempting to plot it using the library's functions, I noticed that the longitude and latitude are reversed (not in the name), and the longitude is plotted on the y-axis. To enhance user-friendliness when using Python tools, it would be beneficial to adjust the format accordingly.

Regarding the availability of MATLAB code, I am not a MATLAB programmer, so I am unable to provide feedback on its use.

Response: This error is caused by the missing value defined as "nan" in the NetCDF file, which can be read by MATLAB but maybe not in other tools. We will replace the missing values to "-999" for all NetCDF files, to ensure compatibility with other tools.