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.

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 traning 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).

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).

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.

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).

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 the total uncertainty of reconstructed pH.

Specific Comments:

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

Lines 45-46: Can be rephrased (for instance: "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".

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

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".

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

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

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!

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

Line 80: "such as the ocean currents product" → should be at 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,...).

Line 116: "Therefor".

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

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.

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!!!

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

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

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.

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

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

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

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 suplot will help to evidence the statements for Fig 5.

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

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).

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

Table 4: report uncertainty estimates for all other products.

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 pCO2 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.