RC2 Comments – PyESPER Manuscript Response

The manuscript describes a new python-based version of the existing ESPER algorithms. No new development or training is performed, but detailed comparison of outcomes with both the original Matlab and the new Python versions is described.

The manuscript is well written and clearly details what is new and how the new version performs compared to the original. It is a nice added value that the algorithms are now available in several programming languages.

I only have some minor comments. Once those issues are fixed I'm happy to see this published.

We thank you for your productive feedback.

Minor issues:

Throughout the manuscript information is needlessly repeated several times. In particular which observational data are included is presented again and again. It should be sufficient to define once what is o and w data, including how much data there is, and then just refer to those definitions. The many repetitions of this information makes the text a bit cumbersome to read.

We have edited the manuscript for repeated information and deleted duplicates of data definitions, including the following:

We removed the (repeated from prior text) number of measurements within GLODAPv2.2022 within the caption of Fig. 1 (L251), definition of open ocean data from captions of Table 1 (L284), Fig. 2 (L297), Fig. 3 (L343), Table 2 (L395), Fig. 7 (L409), and definition of whole ocean data from the captions of Table B1, Fig. B1 (L497), Fig. B2 (L509), Fig. B3 (L515), Table B2 (L524), and Fig. B4 (L537).

L261. Redefinition of open ocean data was removed.

Line 141: ensemble is the more commonly used word so I suggest using only that

The wording has been changed to ensemble, with no mention of committees (now L169).

Line 156-158: The sentence is quite awkwardly phrased. Try revising for clarity.

True. The wording of the entire paragraph has been revised for clarity as follows:

L184-188. The impacts of anthropogenic carbon (C_{ant}) are approximated in ESPER and PyESPERv1.0 using a 1° x 1° gridded transit time distribution (Waugh et al., 2006)-based C_{ant} product referenced to the year 2002 (Lauvset et al., 2016). ESPERs assume that oceanic C_{ant} increases proportionally to atmospheric anthropogenic CO_2 (transient steady state assumptions; Gammon et al., 1982; Gruber et al., 2019; Tanhua et al., 2007). This implies that the "shape" of the C_{ant} vertical profile (gradient) remains constant with continuous exponential increases of atmospheric CO_2 and ocean C_{ant} according to Eq. (3; Carter et al., 2021).

Line 160: I do not understand the meaning of the sentence. Please revise for clarity.

We have reworked the entire paragraph for clarity (see above Line notes).

Line 230-235: All these numbers are also given in the table so it is unnecessary to repeat here. The information is also more easily digestible from a table. Same goes for lines 275-279.

We have replaced these two segments of text with simple reference to Tables as follows:

L280-281. Mean (±standard deviation; $RMSE_n$) PyESPER – ESPER_LIR differences for each property are shown in Table 1.

L338-339. Mean (±standard deviation; $RMSE_n$) offset for each property is shown in Table 2.

Table 2 caption: Try splitting the information into smaller sentences or removing some redundant information.

We have reworked the caption as follows:

L414-417. Table 2: Mean (standard deviation), maximum, minimum, and normalized RMSE ($RMSE_n$) are shown for three scenarios: (1) between Python – MATLAB NNs, (2) MATLAB ESPER_NN – measured values, and (3) PyESPER_NN – measured values. Separate rows exist for TA, DIC, pH_T, phosphate, nitrate, silicate, and oxygen estimates. All units except pH_T are μmol kg^{-1}, and data are for open oceans ($_o$) and all equations combined.

Captions for Figures 2, 7, B1 and B4: There is no information about what the histograms/bars on the top and right represent. This should be added.

We have added the following sentence to each of these figure captions:

Top and bottom side histograms represent the distribution of the x and y axes, respectively.

Lines 287-288: This statement appears to contradict the information on lines 318-319. Please clarify.

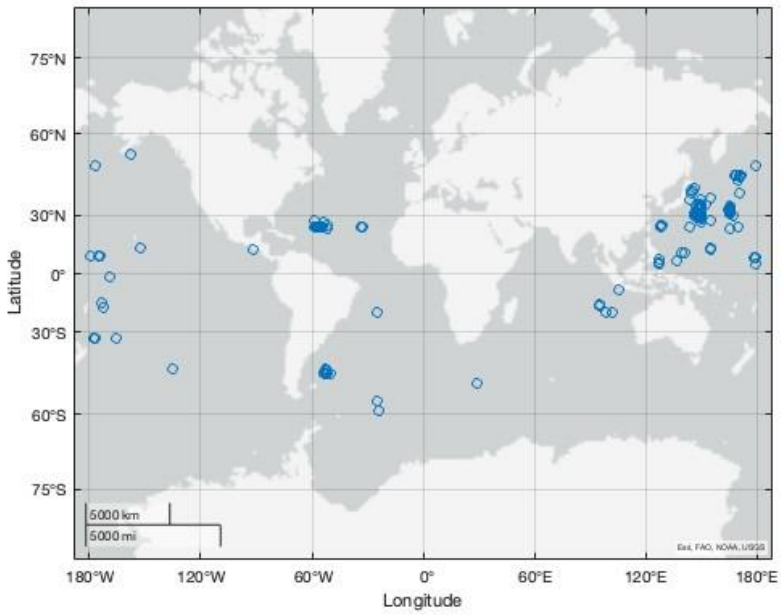
Yes, this is confusingly worded. We have clarified the point we were trying to make, which was that, despite some minor offsets in C_{ant} estimates for pH_T and DIC due to interpolation differences, estimates from NNs for these two variables remain functionally identical. Please refer to the following changes in-text.

L346-347. These minor offsets are attributed to the programming language differences in the interpolation of the C_{ant} adjustment, which is only applied to these two properties.

L395-396. Currently, when C_{ant} estimates are required, the results from PyESPER_NNs remain functionally identical to those from ESPER_NNs, despite minor offsets from the interpolation methods.

Figure 4: Most differences are found in the northwest Pacific Ocean. It would be interesting if you could add a brief discussion about why this is and the implications of it.

The reason for these discrepancies within the western Pacific Ocean is that this is a place where GLODAPv2.2022, which was used for estimate comparisons, contained data from the very deep ocean. Very deep locations are at or near the edge of the original MATLAB grid for training data (5500 m), where interpolation methods had greater differences. You can see the matching "problem areas" on the following (coarse) map of locations where GLODAPv2.2022 samples were collected at >6000 m depth:



We have added the following text to the text on LIR results to better explain this:

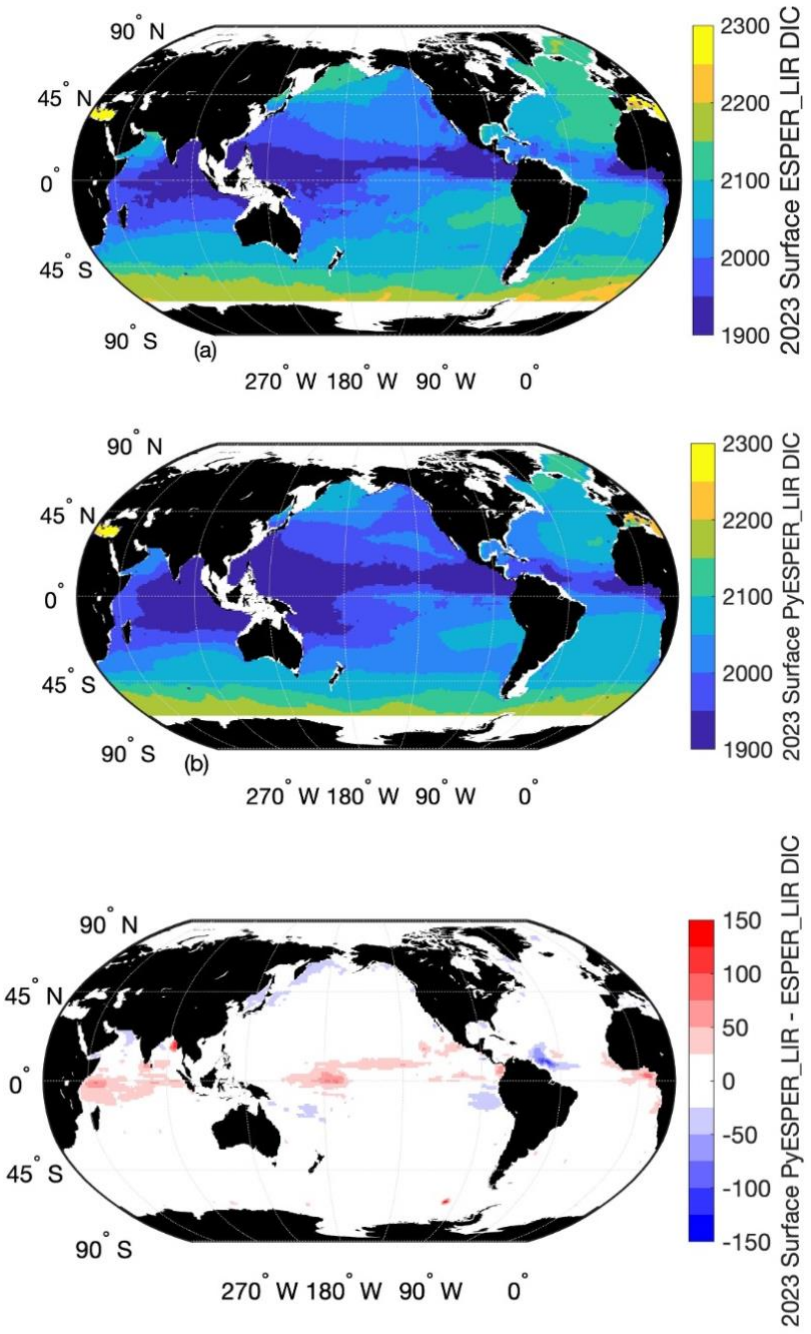
L324-329. PyESPER_LIRs were within 2σ (~95% of measurements should fall within this uncertainty level) for most ocean regions, with a few exceptions which occurred predominantly in coastal areas or deep waters near the edges of the original MATLAB grid (Figs. 3 and 4). Spatial patterns in distribution of outliers shown in Fig. 4 appear to reflect locations where more edge-of-grid biogeochemical measurements were collected (e.g., near coasts and in deep waters). Hence, these locations aligned well with places where coefficients were extrapolated in the MATLAB implementation (see Sect. 2.1.1, "*Locally interpolated regressions*"; Figs. 3, 4, and 5; for $_w$ Fig. B2 and B3).

Line 320: I suggest you rename this section. It is not intuitive that it deals with the differences in speed of calculation

We have renamed the section to "Speed of calculation," as suggested.

Figure 6: It would be useful to have a panel showing the differences between panels a and b

This has been added to Figure 6, as recommended. This figure was moved to appendices (now Fig. C1), as it is likely more suited there. Please see the following modified figure:



Line 358: I suggest to rename the section future work or future improvements.

We have renamed the section "Future improvements"

Please also list the data product doi(s) in the data availability section along with the references.

We have included the doi's for the appropriate data repositories in this section.

Line 383-384: "essentially identical" is not true for DIC and pH. That should be mentioned here too

The estimates are still very closely aligned for DIC and pH_T when estimated using NN methods (C_{ant} contributions account for a slight difference between the MATLAB and Python estimates). We have noted that here and removed the "essentially identical" language as follows.

L489-490. Estimates from PyESPER_NNs precisely align with those from ESPER_NNs for all equations and desired outcome variable combinations (Fig. 7) and estimates from these two routines align very closely for all estimates, and to within machine precision for all but pH_T and DIC, which exhibit slight differences due to impacts of interpolating for C_{ant}.

Table A1: Caption refers to Table S2, but that is really A2.

Thank you for pointing this out. We have fixed this error.