## General comments

Chau et al. present a exclusive approach using discrete ocean surface data of $pCO_2$ and total alkalinity (TA) to obtain a new monthly reconstruction for the period 1985 to 2021 with 0.25º resolution of the marine carbonate system variables. The reconstruction is based on the use of a feed-forward neural network (FFNN) for $pCO_2$. For TA they use locally interpolated alkalinity regression (LIAR). The reconstruction is based on the CMEMS (Copernicus Marine Environment Monitoring Service) product, which provides global reconstructions of sea surface temperature (SST) and surface salinity (SSS) for the same period, including chlorophyll and other physical variables such as sea surface height. The authors start from a previous work where they published a similar database made with a resolution of 1° where only $pCO_2$ has been reconstructed. Here they expand the resolution by increasing it to 0.25º with the inclusion of TA, and then using the thermodynamic equations of the marine carbonate system they obtain the variables: Dissolved Inorganic Carbon (DIC), pH and degree of saturation of aragonite and calcite. In this way a product is generated that can be used to evaluate the impact of ocean acidification by other users and stake-holders. The quality of the reconstruction is contrasted with values observed at a series of oceanic and other coastal time stations. The authors provide two databases, one with 1ºC resolution and the other with 0.25º resolution.

The motivation and idea behind the paper is not original in the sense that this has been done before on a seasonal climate scale, but instead, here, the authors exploit the potential of CMEMS to obtain a reconstruction of all carbonate system variables on a spatial scale that has not been achieved so far and that can certainly be very useful in the evaluation of biogeochemical models and for the study of ocean acidification and in coastal regions of higher variability.

The article is well written and provides detailed information both in the formalization of the equations and in the graphical information that is extended in the figures and equations of the appendix. However, it does not develop a specific discussion section of this new database or a comparison with other climatologies of $pCO_2$, DIC, AT and pH that would allow us to see the benefits, improvements and qualities of the new product. The authors, instead, compare in the 'Conclusions and Discussion' section the acidification rates with other observational results of other authors.

The source of information for the $pCO_2$ reconstruction is the Surface Ocean $CO_2$ Atlas version 2022 (SOCATv2022, 1985-2021) observations of CO2 fugacity ($fCO_2$). This database provides not only $fCO_2$ but the data are **REQUIRED** to be accompanied by SST and SSS. The $fCO_2$ data cannot be used independently of the SST and SSS with which it has been reported, since temperature in the observation of $fCO_2$ has a high impact on the $fCO_2$ value itself (a bias of 1°C generates a bias in $pCO_2$ of 4.2%, ~18 µatm). The development of the $pCO_2$ reconstruction expressed in equation (1) does not meet that requirement. The authors mix the SOCAT observations with the SST and SSS reconstructions of the CMEMS product. This generates important errors as they themselves show in the reconstruction at oceanic (Figure 7) and coastal (Figure 5) fixed stations. Similarly, with TA, the observations used in LIAR also use temperature and salinity in Global Ocean Data Analysis Project bottle data version 2.2022 (GLODAPv2.2022, Lauvset et al., 2022). GLODAPv2 does not report TA without temperature and salinity observations so neither should different data sources be mixed when applying the LIAR methodology as is done in equation 2. Therefore, methodologically, **the manuscript is seriously flawed in its numerical approach**. The process should be done in two stages, first obtaining a set of FFNNs trained with $fCO_2$, SST and SSS with the SOCAT data (and additionally the variables already included in equation 1), and then projecting that FFNN onto CMENS' own reconstructions of SST and SSS. The same is true for TA and the use of LIAR. At least the SSS used in equation 2 should include the GLODLAP SSS and not the CMEMS SSS. Better is to include the GLODAP SST, also. Then the coefficients developed with LIAR are used on the CMENS reconstruction. This would greatly improve the reliability of the algorithms by better reproducing both the oceanic and coastal time series, not to mention that the GLODAP reconstructions shown in Figure 8 will do so as well. All this allows us to have a better estimate of the quality of the obtained algorithms since we can apply them to both oceanic and coastal time series with their own predictors and validate these algorithms. As currently performed in the manuscript, this validation is strongly biased because the SST and SSS reconstructions of CMENS on these series clearly disagree when comparing point data with monthly means as indicated in the manuscript itself in Figure A8.  In addition, a simple linear regression of TA versus salinity would report a better fit than the LIAR model applied in the manuscript.

As shown above, the monthly reconstruction proposed by the authors would be strongly improved if the two-step process is applied. The current product shown has a very poor quality in terms of validity since its comparison with the fixed time-series station used shows very high RMSD values (Figure 5, Table 7 and A3).

**Minor comments**

Line 18: “*reconstructions with root-of-mean–square–deviation from observations less than 8%, 4%, and 1% relative to the global mean*” The relative percentage of RMSD over the mean is not a good parameter to evaluated the goodness of the results. For example, the accuracy of AT is better than 0.1%, and $pCO_2$ is similar. The percentages of RMSD reported are about two orders of magnitude higher. .

Line 20: “*and 0.4% for pH*” It is a bit odd to report percentages of a logarithmic magnitude such as pH.

Line 92: The associated uncertainty reported in the article (σ) refers only to the uncertainty of the 100 replicate FFNNs, but they do not incorporate the uncertainty that each of the FFNNs has with respect to the SOCAT $pCO_2$ values they are trying to replicate. The paper is only assessing a part of the uncertainty, by the way the smallest part and therefore not evaluating the ability of the FFNN set to reconstruct the input values.

Table 1, Table 2 and also Table 3 should include a value or an estimate of the uncertainty of each of the variables, either in their analytical determination or that which each product or reconstruction generates for each of the variables. This helps the reader to evaluate the quality of the reconstruction as a function of own error in the determination of each of the reconstructed variables.

Line 126 Table 3 is cited before Table 2.

Line 214. It is not sufficiently clear how to proceed with the reconstruction. It talks about excluding data in the month of reconstruction. Therefore, it would appear that for each month 100 FFNN reconstructions are performed. If this is correct, the RMSD for each month should be included in the figure or table of the SOCAT $pCO_2$ reconstruction since that data is not used in the month-specific reconstruction.

Line 249 Fig A7 is not cited in order.

Line 285 and 210: How do you solve the discontinuities of the variable 'longitude' around the prime meridian 0°. This is usually solved using the sine and cosine functions of longitude. Any reason for not doing so? Does this variable really bring any improvement in the FFNN?

Table 4. First of all, it should be pointed out that there is an excess of significant figures, not only in this table but throughout the text. Regarding the $pCO_2$ results, the authors should remove all decimal places since analytically its precision is 2 µatm as described in the article. But more importantly, once the superfluous decimal places have been removed, what is observed is that there is practically no significant improvement between the product 'r025' and 'r100'.

Line 350 How is the regriding process performed? What type of interpolation is performed?

Line 354 *'The FFNN(r025) central to this study yields a lower RMSD and a higher correlation to the SOCAT data than the FFNN(r100→ 025)'*. **Unfortunately, there is no significant difference between the two products. This statement is not correct.**

Line 375. The differences in RMSD between the regridded r100 and r025 products are very small, or even in some as in Canary Current System it is larger (strange?). There is no significant improvement in the coastal regions between the two products.

Line 393. It seems a very marginal the 2% improvement in $pCO_2$ reconstruction capability

Lines 404-423. "*Analyzing the eight station time series, we have found that data have been sampled within a few days with an average offset of about a week from the month center. At these coastal sites, the temporal standard deviation from monthly averages of $pCO_2$ exceeds measurement errors (2 µatm, Sutton et al., 2019). pCO2 ranges from 20.12 µatm at GREYREFF to values as large as 65.6 µatm at CAPEARAGO or 69.98 µatm at FIRSTLANDING. The monthly average of $pCO_2$ might not be adequately represented by discreet samples at sites with a large temporal standard deviation of pCO2. The misfit between the monthly reconstruction and discreet observations is exacerbated in dynamical coastal environments and might explain in part the large RMSD of reconstructions of monthly coastal $pCO_2$ (e.g., GREYREEF: 38.34 µatm, CAPEARAGO: 79.86 µatm, FIRSTLANDING: 77.32 µatm) for the r025 reconstruction. The RMSD is mostly lower for the FFNN reconstruction at 0.25º resolution compared to the FFNN at 1º resolution by 2.11 µatm (CCE2) to 23.32 µatm (COASTALMS). Similarly, r2 increases between 7%-23% at higher resolution. Overall, seasonal to interannual variations of coastal-ocean $pCO_2$ are better reproduced in the reconstruction at 0.25º resolution (Fig. 5).*" Here, it becomes evident that comparing monthly reconstructions with point values in coastal areas of high variability results in very low predictive ability on the part of the product produced. As indicated in the general comment, this should be evaluated considering the variability of SST and SSS in the study area because in this way the biases that the CMENS product has to reproduce point values from monthly mean values are being transferred to $pCO^2$. The aforementioned increases in $r^2$ are relatively small if we consider the important biases involved, which in some products even increase as the resolution improves, as in FIRSTLANDING or CHEECAROCKS.

Line 437-438 "*The largest model uncertainty ($\sigma > 30$ µmol $kg^{-1}$) is computed nearshore and surrounding oceanic islands, a feature inherited from input uncertainty associated with the CMEMS salinity product (Fig. A8a)*." This described here is very relevant. In fact, it would be necessary to show graphically the correlation between the uncertainty in TA and SSS in the CMEMS product in both the coastal and oceanic domains. Possibly it shows a very relevant correlation. A similar should be done with the uncertainties of $pCO_2$ and SST in the CMEMS product.

Line 451-465. "*The reconstruction of AT distributions relies on LIAR coefficients fit with GLODAPv2 data (Olsen et al., 2016) covering the years before 2015. These data are also part of the latest version GLODAPv2.2022 (Lauvset et al., 2022). They do therefore not correspond to an independent dataset for the evaluation data of the CMEMS-LSCE reconstruction. To overcome this limitation, reconstructions of AT and DIC are compared to observations for Eulerian time series stations: BATS, DYFAMED, ESTOC, and HOT (see Table 3 and Fig. A1b for data sources and station locations). Figure 7 illustrates the* *comparison between monthly time series of AT and DIC extracted from the CMEMS-LSCE datasets and measurements at these long-term monitoring sites".* These lines and Figure 7 show again how a large part of the discrepancies between the TA and DIC reconstruction is due to the discrepancies in SSS and SST of the CMEMS product, indicating that the reconstruction is not well done. In the case of the DYFAMED station it is very noticeable and contrasts that other products such as climatologies like those cited in the article (Lauvset et al. 2016; Broullón et al. 2019) do not show bias as high as the reconstruction performed here.

Line 473 *"The lowest prediction skill of temporal variability is obtained for ESTOC. Particularly, seasonality to multiyear variations in DIC are predicted at $r^2=0.47$ for ESTOC compared to $r^2 > 0.7$ for BATS and HOT.*" This is not correct. The regression coefficient is not the only criterion for assessing predictive ability. In this case the variability observed at ESTOC is lower than at BATS and HOT, so a lower $r^2$ does not mean lower skill. In fact, the RMSD at is the lowest of all the stations evaluated in TA. In terms of DIC the three stations show similar RMSD.

Line 478 "*Model uncertainty (1σ-envelop) of monthly AT and DIC estimates (Fig. 7a) is also inflated somewhat proportional to the CMEMS salinity product uncertainty (Fig. A10a)."* Evidently. A figure showing that would be useful. That is why including this product in the LIAR training phase for TA does not help to obtain the best possible reconstruction.

Linea 527 "*The reconstructed pH time series reproduce measurement variability with relatively high correlation, $r^2$ in [0.21,0.69], that reinforces the reliability of CMEMS-LSCE pH*". It does not seem that the level of correlation obtained with this reconstruction is significant with such low levels of $r^2$. Additionally, the fact that there is no discussion in the article where these levels are compared with other products even if they are only climatic such as those of Takahashi et al. 2014, or others cited in the article for AT and DIC.

Line 576 .- "Conclusions and Discussion" It should be "Discussion and Conclusions" But on the other hand the **discussion is made not in terms of the assessment of the quality of the reconstruction of the product but in terms of the results in terms of ocean acidification**.

Line 594 "*In comparison to CMEMS-LSCE at monthly and 1º resolutions (Chau et al., 2022b), the reconstructions over coastal areas are improved at higher resolution (Figs. 2-4).*" This is not demonstrated in the article. The reduction in RMSD between the two products is very small or marginal.

Line 609 "*The spatial distribution of long-term mean 1σ-uncertainty estimates (Figs. 1b, 6cd, and 9cd) indicates higher confidence levels for open-ocean estimates than over the coastal sector*". This is very unrepresentative of product quality since it represents the reproducibility of the 100 FFNN but does not evaluate the RMSD between input and reconstructed data.

Table 7 Both $pCO_2$, AT and DIC quantities should not have decimal places (mean, RMSD).

Line 655 No comparisons with other reconstructions like MODO-DIC of Keppler et al. 2020, or AT from Broullon et al. 2019 or Lee et al. 2006.

References: Keppler, L., Landschützer, P., Gruber, N., Lauvset, S. K., & Stemmler, I. (2020). Seasonal carbon dynamics in the near-global ocean. Global Biogeochemical Cycles, 34, e2020GB006571. https://doi.org/10.1029/2020GB006571