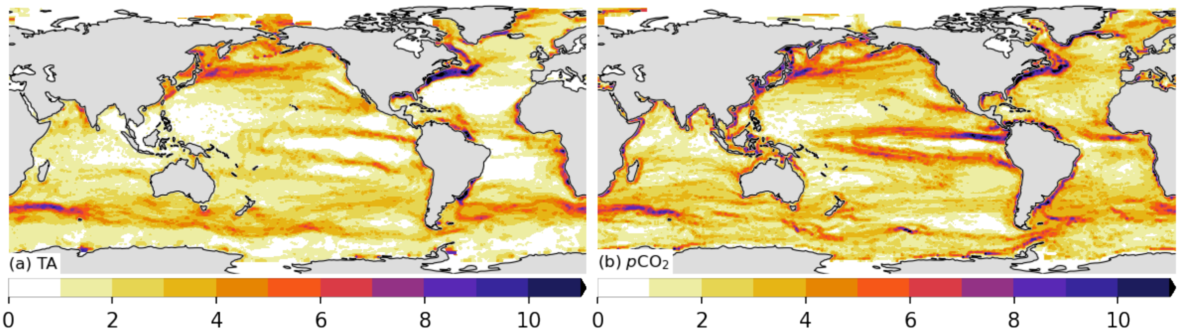
Response to Reviewer 1 (R1)

We thank reviewer 1 for their prompt and positive feedback. Their review was detailed and was a great help in preparing the revised manuscript. Below, we outline how we will address the major points raised : 1) recommendation on how modelers should use the data set; 2) discussion of coastal estimates , particularly considering the new merged Landschützer et al. (2020) product; 3) missing appendix. We included below the reviewer's comments in blue and our responses in black. Italic green font indicates text that will be added to the manuscript. The location of the text in the manuscript is indicated if applicable. The line specific grammar issues will be addressed and shown in the track-changed document. Note that we have also refined our definition of error and uncertainties. These changes have also been marked in the tack changes file.

1. Inclusion of the coastal ocean

It would be good to add 1–2 paragraphs on potential issues in the coastal zone compared to open ocean and recommendations for future efforts. It would also be useful to connect to this recent paper: Landschützer, P., Laruelle, G. G., Roobaert, A., and Regnier, P.: A uniform pCO2 climatology combining open and coastal oceans, Earth Syst. Sci. Data, 12, 2537–2553, https://doi.org/10.5194/essd-12-2537-2020, 2020.

We will add a paragraph in the discussion about the validity of our coastal estimates. In particular, we will outline why we have some confidence in these estimates (with the figure below supporting our point), even though there also some clear limits. Further we will make reference to the MPI-ULB-SOMFFN merged product. Lastly, we will emphasize that users should consult the original SOCAT data for comparison if they would like to use the product beyond the climatological scale. We plan to include the following text in the discussion.



Caption (Appendix): *Map of the position of cluster boundaries across all ensemble members and months for (a) total alkalinity and (b) pCO2. The white regions indicate locations that belong almost exclusively to the same cluster. Dark regions show where cluster boundaries are persistent.*

Discussion: *The OceanSODA-ETHZ product extends further into the coastal margin than most previous studies (Iida et al., 2015; Land-schützer et al., 2016; Denvil-Sommer et al., 2019). This is achieved i) by including coastal observations during the training, and ii) by using a larger number of clusters compared to other clustering approaches (Landschützer et al., 2016; Watson et al., 2020). This permits to better separate open ocean and coastal variability through the inclusion of suitable variables in the clustering step (e.g. Chl-a for pCO_2, and see Figure A3 to see a representation of cluster boundaries). This gives us some* *confidence in the coastal estimates, at least on a climatological scale with regard to the seasonal cycle. Our product is therefore comparable to that of Landschützer et al. (2020) who blended separate coastal and open ocean pCO_2 products into a single climatological product with monthly resolution (Landschützer et al., 2016; Laruelle et al., 2017).*

The total uncertainties of our estimates in the coastal ocean are considerably larger compared to the open ocean estimates (Figure 7). This reflects the much higher spatio-temporal variability of the physical and chemical environment in the coastal ocean, leading to much higher variations in the marine carbonate system (Laruelle et al., 2017). Since our predictor variables are only partially reflecting this variability, a large portion of the high total uncertainty is due to a high representation error (Table 3). Increasing the resolution of the products may improved coastal estimates as done by Laruelle et al. (2017). Until we arrive at this point, the OceanSODA-ETHZ data should be used with care in the coastal ocean. Further, we recommend that researchers interested in the investigation of interannual variability and trends in the coastal ocean using the OceanSODA-ETHZ product should also look a the underlying in situ data to gain a better understanding of the variability, trends, and uncertainties for the coastal region of interest.

2. Recommendation to modelers

You mention ocean models briefly in the introduction (L45). It would be helpful for the community if you could make some recommendations (based on what you have learned in the development of this paper) for accurately simulating and benchmarking ocean acidification in numerical ocean models.

We will add a section with specific recommendations to users of the product, containing a paragraph aimed at the modeling community. Specifically, we will add an additional uncertainty estimate that provides climatologically mapped errors based on test data that has not been seen by the trained model. This will permit modelers to assess model-data misfits on a local basis, thereby making model-data comparisons more quantitative.

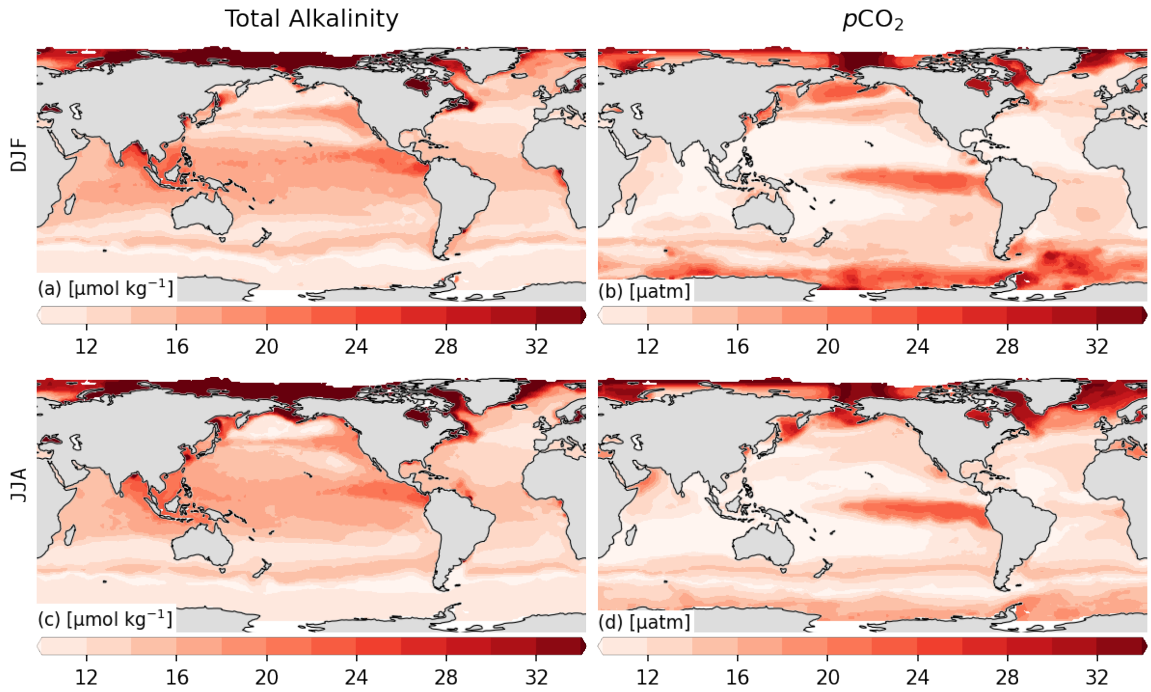


Figure A2: *The Huber test scores mapped to the ensemble clusters for Total Alkalinity (TA) and* pCO_2*. The top row shows Huber scores averaged for December, January, and February (DJF) and the bottom row June, July, and August (JJA). The Huber score is a blend between root mean squared error (RMSE) and mean absolute error (MAE), where MAE is applied to values that are considered outliers. Only test data is used to calculate these climatological scores, meaning that the scores are based on GLODAP2 and SOCAT data for TA and* pCO_2 *respectively.*

Discussion: *The OceanSODA-ETHZ product provides a useful comparison for numerical models with the full marine carbonate system. The spatially resolved climatological estimates of uncertainty for TA and pCO2, based on in situ data, provide useful context for ocean modelers on a climatological time-frame (Figure A2). In the same way that previous studies have used pCO2, the OceanSODA-ETHZ data set can also be used to compare interannual trends and variability of the marine carbonate system (Landschützer et al., 2015, 2016; Gregor et al., 2018; Keppler and Landschützer, 2019).*

(Appendix) *One of the advantages of using the GRaCER approach is that any metric can be mapped from the results to the appropriate clusters, resulting in an ensemble of metric scores. The possible metrics that can be applied include bias, root mean squared error, and mean absolute error. Further, these metrics can be applied to test data, meaning that the resulting scores can be based on test scores — that is data that are unseen by the model during the training process, thus giving a true representation of the uncertainty. Given that the cluster used in this study are climatological, we can get fully mapped climatological estimates of uncertainty.*

Missing appendix

Appendix A3.4 seems to be missing, please correct this

This will be corrected to include text on how the variable importance figure was created. The main text now will refer to the figure instead of the section in the appendix.