Duke et al. Reply on RC2: 'Comment on egusphere-2023-870', Anonymous Referee #2

Summary:

Duke and coauthors use a two-step cluster–regression method to map surface partial pressure of carbon dioxide (pCO2) in the Northeast Pacific Ocean. Their approach is novel in that they grid pCO2 observations and produce pCO2 maps with high spatial resolution (1/12°); in doing so, they offer insightful observations about optimal model parameters and regional driving factors of CO2 flux variability. This work represents not only a useful product for investigating surface carbonate chemistry in the Northeast Pacific (NEP), but a valuable roadmap for increasing the spatial resolution of observation-based surface ocean pCO2 products.

This manuscript is very well-written and clear to follow. I was especially impressed with the analysis surrounding the training of artificial neural networks with progressively finer resolution, and the critical nature of the training/evaluation data split in these instances. The examination of driving factors of CO2 uptake variability and the effects of marine heatwaves is interesting and will be beneficial for researchers seeking a region-wide carbonate chemistry context for the NEP. I detail a few general and line-specific suggestions below, but overall support the acceptance of this manuscript.

Thank you for your time and careful consideration providing feedback on our manuscript. We appreciate your encouragement regarding the potential of our study to serve as a template for global products aiming to achieve higher spatial resolution. Below, we have addressed your comments in a point-by-point manner. Our responses are highlighted in blue, with manuscript text in quotations and added/revised text italicised.

General suggestions:

The conclusion that the training data to internal evaluation data ratio should be optimized and likely increased toward finer resolution grids will be extremely valuable as global-scale observation-based pCO2 data products with finer than 1° resolution are beginning to be produced. In that context, it may be helpful to expand upon the statement at the end of section 3.4 that this result "creates a precedent for stepping to a higher resolution product with nearly no loss in performance". How might you envision that higher resolution step being taken at a global scale? What are some important considerations and potential pitfalls when taking this approach beyond the NEP? Any thoughts about increases to the temporal resolution?

One important consideration with the NEP is that at 1/12° spatial resolution, gridded observation coverage is still actually quite good at 0.39%. In contrast, when looking at the SOCAT global coverage map (https://socat.info/), much of the south Pacific, south Atlantic, and Indian Ocean likely experience a more profound drop off in coverage compared to our Figure 4f "% target pixel coverage" line. Products would be relying on robust nonlinear relationships from the neural network informed in other regions to fill these gaps, raising a flag about critical observational coverage some have covered in the southern ocean (Hauck et al. 2023; https://doi.org/10.1098/rsta.2022.0063). Increasing temporal resolution is difficult and will likely need to be accompanied with changes to how we train the neural network. In increasing temporal resolution, the "% target pixel coverage" line in Figure 4f would move all values closer to zero. We feel the predictor training data may need to shift to using *in situ* data from high frequency underway systems prior to gridding to establish nonlinear relationships in the neural network, before labeling higher temporal resolution gridded predictor data with estimated pCO_2 values. This approach could mark a major shift in the practice of creating observation-based products and significantly increase in computing costs.

Considering this discussion, the following text has been added to section 3.4 line 280.

"In regions with sufficient observational coverage (Figure 4f; Bakker et al., 2016), this finding creates a precedent for stepping to a higher resolution product with nearly no loss in performance, overcoming the overfitting concern with increased resolution (Rosenthal, 2016)."

One limitation of the validation performed here is that the statistical metrics represent the ability of the ANN-NEP procedure to estimate pCO2 only at the spatiotemporal grid cells where observations are available. This may mask location-specific seasonal biases, especially at high latitudes where wintertime observations are likely not as plentiful. In lieu of a comprehensive model simulation experiment to evaluate these unquantified biases, this consideration may warrant some discussion in section 3.2 or elsewhere.

RC2 is correct. We added the following text to section 3.2 line 225.

"One limitation of our approach in assessing the uncertainty of the ANN interpolation method is that it is only applicable to grid cells where observations are available. Consequently, location-specific seasonal biases, especially in high latitudes with limited wintertime observations (Figure 1a&b), may not be fully captured or accounted for."

A figure displaying the most frequent occurrence of each SOM province over the timeseries would be informative. As an additional suggestion for future work: to reduce the discontinuities at the borders of biogeochemical provinces it would be interesting to explore soft clustering approaches in addition to hard clustering like SOMs. Soft clustering approaches provide probabilities for each clustered grid cell, which can be used as weights to blend pCO2 predictions across clusters.

A figure displaying the mode SOM biogeochemical province in each pixel has been added to the supplementary and in text citation section 2.4 line 139.

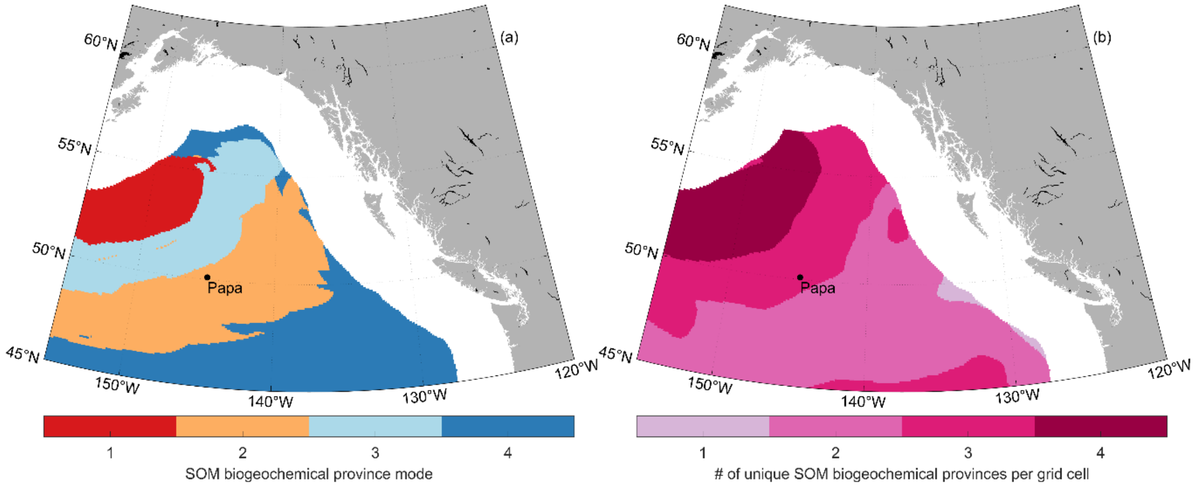

Figure S2 Mapped (a) mode of SOM biogeochemical provinces (i.e., most frequent occurrence), and (b) the number of unique SOM biogeochemical provinces each pixel belongs to for each month from January 1998 to December 2019.

Moving to a soft clustering approach to remove artificial fronts created by the province boundaries is a great suggestion for future work.

Line-by-line comments:

Line 85: It would be valuable to articulate why the coastal ocean was excluded in this study.

The coastal ocean experiences much greater variability and presents all sorts of unique challenges. To create a "good" coastal product for the region, we felt that specific tuning measures would be needed. These issues will be addressed in a separate stand-alone paper led by PD – and is currently in progress. We have added to the text in section 2 line 85:

"We limit our study region to the open ocean regions with reduced variability and related drivers compared to the continental shelf regions. Creating a product on the continental shelf and in the nearshore requires different neural network considerations and is associated with high uncertainties (Roobaert et al. 2023)."

Lines 142–143: It isn't immediately clear why choosing not to normalize predictor data implicitly weights the SOM predictors toward the pCO2 climatology. Is it related to the relative range of each chosen predictor?

Yes, in not normalizing the SOM predictors data we forced the relative weights of the input data toward the pCO2 climatology, as the range between the lowest and highest value of pCO2 is at least one order of magnitude larger than that for SST, SSS, and log(MLD).

Climatological SOM predictor variable	**Mean ± standard deviation**	**Range**
Climatological sea surface pCO_2 (μatm)	355±18	209

Sea surface temperature (SST; °C)	10±3	14
Sea surface salinity (SSS)	32.6±0.1	0.7
Log mixed layer depth (log(MLD))	3±1	2

We added the following text to section 2.4 line 143.

"We did not normalize predictor data (e.g., force a mean of 0 and standard deviation of 1), implicitly weighting SOM predictors toward the pCO2 climatology *as its range is at east one order of magnitude greater than that of SST, SSS, and log(MLD)* (Landschützer et al. 2013)."

Lines 151–152: I don't understand what is meant by "we introduced each predictor variable again after deseasonalizing". Can this be explained more clearly?

In total we use 12 predictors in the FFN regression step being all those in table 1 plus those in table 1 deseasonalized:

1. Sea surface temperature (SST)
2. Chlorophyll-a (Chl)
3. Sea surface salinity (SSS)
4. Sea surface height (SSH)
5. Mixed layer depth (MLD)
6. Atmospheric pCO2
7. Sea surface temperature anomaly (SST)
8. Chlorophyll-a anomaly (Chl)
9. Sea surface salinity anomaly (SSS)
10. Sea surface height anomaly (SSH)
11. Mixed layer depth anomaly (MLD)
12. Atmospheric pCO2 anomaly

Update the text to include:

"To emphasize interannual and longer-term trends within the six predictor variables (Table 1), *each predictor variable is used in two different forms, first in its raw form and second after deseasonalizing, bringing the total number of FFN predictors used to 12."*

Lines 280–281: Very interesting and insightful conclusion!

Thank you.

Lines 455–456: It would be good just to clarify in this sentence that "stepping to a significantly higher spatial resolution" refers to a higher resolution "than typical observation-based pCO2 products (1/4° or 1° resolution)" or something along those lines.

Revised the text to include:

"We found that stepping to a significantly higher spatial resolution*, compared to typical open ocean observation-based pCO2 products (1/4° or 1° spatial resolution),* led to nearly no loss in performance despite a much lower ratio of gridded pCO2 observations compared to the total number of grid cells."