Earth Syst. Sci. Data Discuss., referee comment RC2
https://doi.org/10.5194/essd-2021-328-RC2, 2022

Reviewer Comment on essd-2021-328

Anonymous Referee #2

Referee comment on "A monthly surface pCO_2 product for the California Current Large Marine Ecosystem" by Jonathan D. Sharp et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2021-328-RC2, 2022

A review of

A monthly surface pCO2 product for the California Current Large Marine Ecosystem

J. D. Sharp et al.

General Comments

Sharp and coauthors produce a pCO2 observation-based product for the California Current region, expanding and improving upon available coastal products. Their monthly product extends back to 1998 and is at a 0.25 degree spatial resolution and utilizes a Random Forest Regression machine learning technique. Comparisons to other available products as well as in situ observations are thorough and clear.

The paper is clearly written with an important product for the ocean carbon community. I have a few simple questions to clarify and suggestions for improvement of the flow of the text, but other than that support publication of this description of the available product.

Specific Comments

Overall I feel that the abstract could be improved. For one, it doesn’t include the years that this new product (currently) covers. In my opinion, the first half of the abstract reads more like an introductory section and did not give me the information about the excellent

product available, the specific improvements it exhibits over other available products, and it's potential uses for the community. I suggest editing to expand on those important aspect further.

Also, the sentence that begins on line 36 seems disjointed from the topic of the previous sentences: anthropogenic CO_2, impact of dissolving CO_2 on ocean life, etc which then jumps to calculating CO_2 transferred between ocean and atmosphere needs the partial pressure. Consider revising to improve flow of this first paragraph.

In the discussion of how the gridded observations is created for this product (Line 135), I would be interested to see how the mean and standard deviation values vary between platforms. Could that be included in the manuscript or is it available through SOCAT somewhere?

The discussion of your machine learning method is thorough and clear. I find the discussion of three strategies for assessing skill of your model (Section 2.6) very interesting and specifically find the third strategy most exciting as local seasonality is a topic tough to dig into with coastal locations where the seasonality would be enhanced but observations tough to come by. I suggest highlighting the findings related to these tests. I find the results shown in Figure 2 (and associated supplementary figures) to be the most striking and perhaps the most clear take-away from the paper.

For Figure 4, is the N value (i.e. the number of dots on each plot) the same and could you include that number on the plot. These types of plots are sometimes hard to glean that information from since so many of the points are overlaid. From its current presentation it doesn't look like the N value would be the same for all 4 subplots.

For Figure 5, the moorings line has the largest standard deviation/shading. I was wondering if that uncertainty bound is also calculated from the monthly mean values, as would be the case for the products being shown as comparisons (for L20, L17, and RFR-CCS-clim). I wanted to ensure it wasn't simply that difference that is causing the larger spread in the moorings observations vs the products.

Section 3.5 has a clear discussion of uncertainty calculations and the three main sources of uncertainty considered by the authors. With a stated uncertainty of 43.6uatm in the coastal regions for this product, how do the authors suggest users go forward with that information and utilize this product given that large uncertainty range? Also, this ties into the Figure 8a plot where no uncertainty range is provided for the RFR-CCS seasonal cycle. I would imagine that given an uncertainty that large, that the spread on the resulting flux estimates would be significant. I would suggest changing the plot to include the individual estimates from the products included in the SeaFlux product (rather than the shading showing the range) and then include a shading for the uncertainty associated with the RFR-CCS product. Obviously, each of the products included in SeaFlux would have their own associated uncertainties, similar to those described here for RFR-CCS, but would be good

to see regardless. Overall though, I find the comparisons discussed in Section 3.7 to be strong, clear, and decisive.

Overall/Next step Comments

I finished the paper wishing that there was more discussion on how to extrapolate this work to all coastal regions, or at least eastern upwelling regions with similar physics. Is there "sufficient" data anywhere else in the coastal regions to do such a product? Would the authors expect that the relationships from the RFR would be similar in other coastal regimes? Or would other machine learning methods perhaps be better utilized in areas with less data density? Your conclusions does touch on this somewhat, suggesting that the neural network approach use an increased number of clusters for the coastal ocean areas, but I wonder if you could comment more on this. Could SeaFlux be updated with this (and other) regional coastal products to help fill in missing areas omitted by the open ocean products? Or will the open-ocean products gradually start expanding their coverage to coastal regions with improved machine learning techniques catered specifically to coastal as you have done here?