Earth Syst. Sci. Data Discuss., referee comment RC3
https://doi.org/10.5194/essd-2022-71-RC3, 2022

Comment on essd-2022-71

Anonymous Referee #2

Referee comment on "A new estimate of oceanic CO_2 fluxes by machine learning reveals the impact of CO_2 trends in different methods" by Jiye Zeng et al., Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2022-71-RC3, 2022

Review of

A new estimate of oceanic CO2 fluxes by machine learning reveals the impact of CO2 trends in different methods

By Zeng et al

Summary

The authors present an updated pCO2 data-based machine learning product which specifically focuses on improving the rate of increase estimated by the model approach. Rather than a constant trend correction to collapse available data to a base-year for analysis, they use a variable rate. Further comparisons with other DML approaches which similarly use a fixed trend reflects that the ocean carbon flux estimate would be overestimated at both ends of the time series with such an approach.

While this finding is interesting and important for the community overall, this paper suffers from broad statements that are not supported with literature citing or sufficient analysis as well as numerous grammatical errors. While the product itself is a valuable update/addition to the field, revisions to this manuscript should be required before acceptance.

Broad statements

The explanation of the creation of the product could be improved to make it clearer. Specifically, the choice of outlining the steps 2.1 and 2.3. Additionally, comparisons in figures are made as compared to a fixed-rate estimate, but I would be interested to see comparisons made to the same method if a simple atmospheric xco2 time-varying trend was used instead. Would the improvement be seen with that choice and the rate extraction step not providing much benefit? There isn't a clear comparison of that and from Figure 1a it seems like it would do pretty well with atmospheric xCO_2.

I feel that there are many, many places in this manuscript where the statements are too broad and unclear or over-ambitious. In the abstract itself, where it states, "the result could be largely affected by how ocean CO2 trends are obtained", this doesn't make sense because other DML techniques are not assigning or removing "ocean CO2 trends". They are using atmospheric xCO_2 as one (of many) driver variables and the machine learning approach estimates the pCO_2 in areas missing data and from that ocean trends can be calculated. Therefore, this type of statement is not appropriate and especially not in the abstract.

Additionally, the statement in the abstract, "…and the sink in early and late year of the modelled period could be overestimates if CO2 trends were not well processed by models" is incorrect. Yes, this product has a smaller sink as compared to the others, but it does not show a difference in both early and late periods as compared to all other products- just those that use a fixed rate to correct observations to a base year. And what is "not well processed by models"? You must be careful about your terminology. Models are different from DML products as discussed in this and similar literature. So, to say that it is not well processed by models just doesn't make any sense. The models are not processing the data- the machine learning methods are.

Regarding the comparison to other products, specifically Figure 6, first, the layout seems like a waste of space for these figures and it could be condensed more (also include a legend on each figure subplot for ease of interpretation). Second, the authors leave out a comparison to the JENA product which is one of the most long-available pCO_2 products and has been included in the GCB for many years now. And lastly, they don't account for the fact that the products all are calculating flux using their own choice of wind and atmospheric products, gas exchange parameterization and other choices. These choices could easily make the difference in the increase in flux seen in the last decade or so as compared to the presented NIES -ML3 product. For example, if the wind product chosen by CMEMS-FNN has stronger winds in the 2000s, it could result in a larger flux and therefore that would account for some of the difference seen here. Unless you are standardizing for all of these factors (as done in the pySeaFlux package/product ensemble) then you cannot confidently say the difference are due to their choice of method regarding co2 trend handling.

Another overarching question I had when reading through this method is that the authors need to be sure to adequately handle the fact that global location of the observations are totally variable in time and that definitely impacts the trend. There is no way to account for or correct for the fact that some years data is dominated from the tropical pacific which could have a much different trend than that of the rest of the ocean. This could play a large role in the differences seen as well.

The effort that the authors put to explain their choice of rate and how it impacts the model result is substantial and accounts for about half of the manuscript. But in the end, they get a value that is very close to the (smoothed) atmospheric xco2 increasing rate. The comparisons made in Figure 2 are great at showing a single trend is inappropriate for such a long-time scale of data that is now available, but it doesn't show why this method of rate assessment is an improved method over just using atmospheric xCO_2 trends.

In Section 3.2 where the flux is presented, the authors report an uncertainty value. It is important to note that this uncertainty is solely due to the machine learning choice used, and represents a spread around the three methods shown here. There is no uncertainty included from other aspects of the flux calculation such as the flux parameterization for example.

The flux convention used in this manuscript is opposite of that typically used in the observation-based pCO_2 product literature. First and formost though, the convention is never defined and explained that a negative flux value would mean an efflux of carbon out of the ocean. I am confused as to why the authors chose this convention in the first place, especially then used a colorbar that is opposite of expected (negative values in warm colors) so that the maps look like the maps in the literature presenting other products. I would highly suggest flipping the convention around, specifically for the maps and discussion in section 3.2.

In regard to Figure 6 and the comparison to other available DML products, it could be that the differences seen in the later years could be influenced by the nearly logarithmic increase in available data for recent decades as opposed to the 1980s and 1990s. I would be interested to hear the authors comment son this in the discussion in Section 3.3.

Specific comments

There are way too many grammatical and clarifying errors to correct and list here. Some common ones are the use of the word "residue" where it likely should be "residual" and "special" where they meant "spatial".

Line 149: “The YON was about 20 years in early 1990s” as shown in Figure 1b- I don’t see that from Figure 1b at all. It seems that the line has flattened well before 20 years on the x-axis. What makes the author state this value here?

Line 152: An example of where the authors are over-confident in their statements. “One of the reasons must be that the data points after 2000….” Is that really the ONLY reason that could possibly explain this? I don’t think so. A better phrase would be that one of the reasons “could be”…

Line 194: Figure 3 shows the annual fluxes and Figure 4 shows the spatial distributions.

Line 195: Typo- should be Global Carbon Budget 2021.

Figure 4: Include an explanation of what a, b, and c are in the figure caption. If you are going to state that “the patterns agree well with those in GCB-2021, then show this comparison. What models is this comparing to, and which compare well? The models, spatially, actually have a LARGE range of estimates of flux (and thus pCO2), see Fay & McKinley 2021 in GRL. To broadly say that these maps compare well with the GCB2021 models is untrue and unsupported in this paper.

Line 207: Why are you comparing your 1980-2020 to GCB-2021 for years 2011-2020? You show in Figure 6 that these are years where many of the products start to diverge.

Why not do the comparison for the same subset of years?

Line 560: “GSB-2021” should be “GCB-2021”.